# Estimating the U.S. Citizen Voting-Age Population (CVAP) Using Blended Survey Data, Administrative Record Data, and Modeling Technical Report

**by**

**J. David Brown**
**Genevieve Denoeux**
**Misty L. Heggeness**
**Carl Lieberman**
**Lauren Medina**
**Marta Murray-Close**
**Danielle H. Sandler**
**Joseph L. Schafer**
**Matthew Spence**
**Lawrence Warren**
**Moises Yi**

**U.S. Census Bureau**

**CES 23-21** **April 2023**

**Abstract**

This report develops a method using administrative records (AR) to fill in responses for nonresponding American Community Survey (ACS) housing units rather than adjusting survey weights to account for selection of a subset of nonresponding housing units for follow-up interviews and for nonresponse bias. The method also inserts AR and modeling in place of edits and imputations for ACS survey citizenship item nonresponses. We produce Citizen Voting-Age Population (CVAP) tabulations using this enhanced CVAP method and compare them to published estimates. The enhanced CVAP method produces a 0.74 percentage point lower citizen share, and it is 3.05 percentage points lower for voting-age Hispanics. The latter result can be partly explained by omissions of voting-age Hispanic noncitizens with unknown legal status from ACS household responses. Weight adjustments may be less effective at addressing nonresponse bias under those conditions.

**Keyword*:*** citizenship, administrative records, voting-age population, nonresponse bias

**JEL Classification:** J1, C1, C6, C8

* Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed, with Disclosure Review Board Release Number CBDRB-FY23-272. This work would not be possible without extensive support of U.S. Census Bureau leadership including: Deputy Director Ron S. Jarmin, Associate Director John M. Abowd, Associate Director Victoria A. Velkoff, and American Community Survey Office Chief Donna M. Daily. We are thankful to additional leaders and staff who contributed to this project, including Mark Asiala, Adam Bee, Gary Benedetto, Edward Castro, Leah Clark, Stephanie Coffey, Suzanne Dorinski, Jonathan Eggleston, Amanda Eng, Michael Freiman, Kendall Houghton, Brian Knop, Erin Love, Linden McBride, Andres Mira, James Noon, Kate Pennington, Maria Perez Patron, Nikolas Pharris-Ciurej, David Raglin, Rolando Rodriguez, Jonathan Rothbaum, Cameron Scalera, Mehrgol Tiv, Evan S. Totty, Lee Tucker, Erik Vikstrom, Caroline Walker, James Whitehorne, Jennifer Withrow, and Tommy Wright.

**Estimating the U.S. Citizen Voting-Age Population (CVAP) Using Blended Survey Data, Administrative Record Data, and Modeling**

**Table of Contents**

# Abbreviations

ACS = American Community Survey

AIAN = American Indian and Alaska Native

AR = Administrative Records

BOP = Federal Bureau of Prisons

BW = Initial ACS housing unit base weights

CAPI = Computer-Assisted Personal Interview

CHCK = Census Household Composition Key file

CMID = Continuous Measurement Identification number (ACS housing unit sample case)

COVID-19 = Coronavirus disease of 2019

CPS = Current Population Survey

CVAP = Citizen Voting-Age Population

DMS = Data Management System

DRB = Disclosure Review Board

EBW = Entropy Balance Weights

GQ = Group Quarters

HU = Housing Unit

IRS = Internal Revenue Service

ITIN = Individual Taxpayer Identification number

MAF = Master Address File

MAFID = Master Address File Identification number

MOE = Margin of Error

NACS = PIK associated with ACS HU, but not in ACS household roster

NAR = No Administrative Record Citizenship

NCOA = U.S. Postal Service National Change of Address file

NH = Non-Hispanic

NHPI = Native Hawaiian or Pacific Islander

NPVS = Not Sent to PVS Search due to confidentiality concerns

NRFU = Nonresponse Followup

NSS = Not Sent to PVS Search

NUMIDENT = Numerical Identification File

OLS = Ordinary Least Squares

OMB = Office of Management and Budget

PEP = Population Estimates Program

PES = Post-Enumeration Survey

PII = Personally Identifiable Information

PIK = Protected Identification Key

PNUM = Person Number in ACS

PPM = Person-Place Model

PVS = Person Identification Validation System

SIPP = Survey of Income and Program Participation

SNAP = Supplemental Nutrition Assistance Program

SS = Sent to PVS Search

SSA = Social Security Administration

SSN = Social Security Number

TANF = Temporary Assistance for Needy Families

TW = Traditional Weights (used prior to 2020)

USCIS = United States Citizenship and Immigration Services

USMS = United States Marshals Service

# Executive summary

Since 2011, the U.S. Census Bureau has produced annual special tabulations of the Citizen Voting-Age Population (CVAP) based on responses from the 5-year American Community Survey (ACS). These tabulations are developed at the request of the U.S. Department of Justice to support its enforcement of Section 2 of the Voting Rights Act of 1965.

The emergence of the global COVID-19 pandemic in early 2020 caused major disruptions to data collection and other survey operations of the ACS, leading to a 2020 housing unit response rate of 71.2 percent, much lower than in earlier years (94.7, 93.7, 92.0, and 86.0 percent in 2016, 2017, 2018, and 2019, respectively).[2] The standard set of data products was not released.

For the 2016-2020 5-year ACS estimates, the Census Bureau attempted to remediate the additional 2020 nonresponse and coverage error by using a modified set of survey weights, called entropy balance weights, in the processing of the 2020 data. The entropy balance weights were informed by records from other sources, called administrative records (AR), and survey auxiliary data comparing characteristics of responding and nonresponding housing units (Rothbaum et al. 2021).

In this report we discuss the development of a method using AR to fill in missing values resulting from questionnaire item and housing unit nonresponses. This work was commissioned when it became clear that the 1-year ACS data did not meet statistical quality standards. The CVAP special tabulations use 5-year ACS data to support tabulations at lower levels of geography, but it was unclear when this project started whether the traditional 2016-2020 5-year ACS data would meet quality standards and be released as planned. The Census Bureau attempted to remediate the additional 2020 nonresponse and coverage error by using a modified set of survey weights, called entropy balance weights, in the processing of the 2020 data. The entropy balance weights were informed by records from other sources, called administrative records (AR), and survey auxiliary data comparing characteristics of responding and nonresponding housing units (Rothbaum et al. 2021). Though the 5-year ACS using traditional weights in 2016-2019 and weights incorporating entropy balance weights in the 2020 data processing ultimately met statistical quality standards, and the CVAP special tabulations using those data were released, we investigate new uses of AR to address survey nonresponse bias, which was present before the pandemic and is likely to persist. This work supports the Census Bureau's ongoing efforts to make its major data products more robust to catastrophic interruptions of fieldwork and long-term trends in willingness to respond to surveys.

## *Enhanced CVAP Method*

This project's enhanced CVAP method blends ACS and AR data in the following way. The record linkage used to combine ACS and AR data comes from the Person Identification Validation System (PVS), which assigns Protected Identification Keys (PIKs) to person records that can be linked to reference files consisting of government records. Citizenship values reported in the ACS and unaltered in the editing process (i.e., as-reported values) are used when available. AR citizenship is constructed using a set of business rules applied to the following data sets: the Social Security Administration (SSA) Numerical Identification file (NUMIDENT); U.S. passports; U.S. Citizenship and Immigration Services (USCIS)

[2] See https://census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/.

naturalization, lawful permanent resident, refugee, and asylee data; Individual Taxpayer Identification numbers (ITINs); Federal Bureau of Prisons; U.S. Marshals Service; state driver's licenses; state public assistance programs[3]; and as-reported values from other Census Bureau household surveys.

In the enhanced method, all ACS data provided by the respondent in the 5-year ACS file are used including as-reported citizenship, but ACS values for citizenship that come from the imputation, assignment, and allocation processes are not used. When the ACS does not have an as-reported value for citizenship, this enhanced method provides one as follows. If a value for citizenship is available in the AR, that value is used. If no AR citizenship is available, we divide these people into separate groups by the reason they are missing AR citizenship (e.g., the ACS person record has insufficient personally identifiable information (PII) to receive a PIK vs. has sufficient information but could not be linked to the PVS reference files). Note that citizen shares vary significantly across these groups as well as among people with as-reported ACS citizenship. We then estimate logistic regression models with training samples consisting of subsets of each of those groups. As-reported ACS citizenship is the dependent variable in these regressions. The probabilities generated from this process are then used for the citizenship values. We used this method to enhance citizenship values for people in ACS responses lacking an as-reported citizenship value. However, there were also households where the entire household unit did not respond to any questions on the ACS questionnaire. Our method for these households is described below.

To create age, race, and citizenship values for people in a nonresponding ACS housing unit (unit nonresponse), we construct a set of people with PIKs with at least one AR source listing that ACS address (MAFID) within a year of the original ACS tabulation month. We predict the probability that the person resides at the ACS housing unit using a logistic regression model trained on the ACS data. This probability is normalized so that the sum of the person's probabilities of residing at their different AR addresses within a year of the ACS tabulation month equals one. When available, we attach age from the NUMIDENT to these people. Race/ethnicity comes from several survey and AR sources (when available), with a set of business rules adjudicating any discrepancies. Otherwise, we use predicted probabilities for age and race/ethnicity. Age (voting age (age 18 and above) vs. non-voting-age (0-17)) is predicted using a logistic regression, and race/ethnicity is predicted via a multinomial logistic regression model. Citizenship comes from AR as described above, when available. Otherwise, we use predictions from a logistic regression model trained on people with as-reported ACS citizenship who have a PIK but do not have AR citizenship, the group our research suggests best represents this component of the population.

As mentioned above, there is no replacement of as-reported ACS citizenship with modeled or administrative data. The probabilities for all ACS responses other than citizenship, including person-place probabilities for ACS household roster members, age, race/ethnicity, are forced to equal one, as is the case for as-reported ACS citizenship. After these ACS and AR data are combined using the methodology above, disclosure avoidance procedures are applied via swapping using the current ACS swapping criteria modified to accommodate all units in the frame.

[3] The state public assistance programs include Temporary Assistance for Needy Families (TANF) and Supplemental Nutrition Assistance Program (SNAP). We have access to data from a subset of the states, only some of which have citizenship data.

We then tabulate citizenship by age by race/ethnicity using the completed, swapped data for all ACS sampled units.[4] A person's weight in each cell is based on the probabilities that the person resides at the housing unit in the ACS tabulation month and of having the cell's citizenship, age, and race/ethnicity characteristics, their housing unit's ACS base (initial sample) weight, and a population control adjustment.[5] The population control adjustment uses the average of the 2016-2020 Population Estimates Program (PEP) county-level population by age and race/ethnicity. To calculate margins of error, the tabulation is conducted 80 times using separate replicate base weights and their respective population control adjustments. ACS special tabulation rounding rules are applied to the final estimates for additional disclosure protection.

## *Estimates Comparison*

We compare the enhanced CVAP estimates to those using only ACS response data weighted by traditional ACS person weights in all years as well as those including entropy balance weights in the processing of the 2020 data.[6] We also compare total population and population by age and race/ethnicity with the 2020 Census values in the initial 2020 ACS housing unit sample. The main results are as follows.

- Modifying person weights to incorporate AR and other auxiliary information about differences between responding and nonresponding housing units has almost no effect on the estimates.
- The enhanced CVAP method produces a citizen share estimate 0.74 percentage points lower than the final ACS-based estimates that were released.
- Using AR to fill in data for nonresponding units has a larger effect on the citizen share than inserting AR and modeled citizenship in place of edited and imputed citizenship in ACS responses.
- The voting-age Hispanic citizen share is sensitive to the choice of method. It is 3.05 percentage points lower when using the enhanced CVAP method than when using the final 5-year ACS official data.
- The final 5-year ACS produces closer total population estimates to those in the PEP, while the enhanced CVAP produces closer age-race/ethnicity estimates.
- As-reported ACS and AR citizenship have a very high agreement rate (99.54 percent).
- Modeled citizenship estimates more closely agree with as-reported ACS citizenship patterns across groups with different reasons for not having AR citizenship than production ACS imputations do.

---

[4] To protect the confidentiality of ACS survey responses, the Census Bureau applies a procedure called "swapping" before releasing data to the public. Swapping switches the data from a number of ACS households with the data from households with similar characteristics in different locations. The pre-disclosure (unswapped) ACS data files are internal Census Bureau files that have not undergone swapping.

[5] All the probabilities are unconditional.

[6] We use the term "traditional weights" to refer to weights constructed using the ACS weighting methodology used prior to 2020. In one version of the ACS used here, the traditional weights are applied to all years between 2016-2020. "Entropy balance weights" refers to traditional weights applied to 2016-2019 and final weights incorporating entropy balance weights in the processing of the 2020 data. Additional processing was applied to the entropy balance weights to produce the final weights for the 2020 data. This latter version was used for the 2016-2020 5-year ACS and CVAP releases.

- Production ACS edits and enhanced CVAP modeled citizenship produce similar estimates for item missing data.
- Using AR to construct household rosters produces population estimates similar to the 2020 Census.
- AR household rosters and the 2020 Census values for the same housing units show very similar differences in the age and race/ethnicity distributions between responding and nonresponding ACS housing units.
- Two factors, in relatively equal proportions, can explain most of the overall difference in citizen shares across methods. One is using person data in nonresponding units instead of performing a probability of selection adjustment and a nonresponse adjustment to responding units.[7] The second is using administrative records instead of a survey data collection mode to produce household rosters.[8]
- Among voting-age Hispanics, using administrative records rather than survey data to construct household rosters explains most of the difference between the ACS and AR citizen share. This appears to be driven by high omission rates of voting-age Hispanic noncitizens from ACS responses, especially those with unknown legal status.
- Population control weight adjustments have little effect on the enhanced CVAP estimates, reflecting the similarity between the blended ACS-AR data without population controls and PEP estimates.
- Disclosure avoidance protection via swapping has little effect on the coefficients of variation of the enhanced CVAP estimates.

## *Potential Improvements to Enhanced CVAP Method*

Two measures could expand coverage of the noncitizen population. One is to use an expanded set of reference files that include people without SSNs or ITINs. A second is to broaden the set of AR sources that cover noncitizens. These changes would likely lower the enhanced CVAP estimated citizen share.

Gaining access to AR data that inform whether a person is a U.S. resident or not at any given time would be beneficial to model this probability. Missing U.S. resident status is a significant source of uncertainty about estimates based on AR data. Incorporating this probability would probably increase the enhanced CVAP estimated citizen share.

Some of the enhanced CVAP design choices could be revisited. The comparison of ACS to AR household rosters suggests that the ACS may be omitting noncitizens at a higher rate than citizens. Besides the option of using AR households in place of ACS households, another option would be to blend the ACS and AR household roster information. For example, if a person-place model predicts that an AR person is likely a resident in a household, but (s)he is not in the ACS response, then the AR person could be

[7] Among housing units that do not self-respond, a random sample is selected for CAPI follow-up, and they remain in the sample. The non-selected cases are subsampled out. A probability of selection adjustment is applied to the weights for the cases that are selected to remain in the sample.

[8] The administrative record vs. survey comparison is how administrative records differ from what the survey data would have been if the nonresponding households had responded.

added to the response. Likewise, probabilities of being a resident of the housing unit in the ACS tabulation month could be applied to the ACS household roster members as well.

The enhanced CVAP uses only addresses selected for the ACS from its address frame. An alternative would be to include all U.S. addresses with people in the AR. Using the full AR population would improve accuracy and reduce standard errors at lower levels of geography. The estimates would also be timelier since it would not be necessary to combine multiple years of data.

# 1 Introduction

Survey nonresponse bias is an increasing concern in statistical and social science communities, especially since the advent of the COVID-19 pandemic,[9] but the issue is not new. Decreasing response rates both at the household level (unit nonresponse) and the individual question level (item nonresponse) have been documented throughout the past decade in major U.S. household surveys. Several studies have reported increased item nonresponse of income, earnings, program participation, and other sensitive questions (Bollinger et al. (2021); Brown et al. (2019a); Meyer and Mittag (2019)), and the use of administrative records has been suggested to improve survey quality (Davern, Meyer, and Mittag (2019)).

In the past, Census Bureau statisticians addressed item nonresponse through hot-deck-like imputation procedures and unit nonresponse through weighting adjustments. Taking advantage of increasingly available administrative records (AR), the Census Bureau has been investigating the direct use of AR to fill in survey item nonresponses (e.g., Brummet et al. (2018)), as well as to inform survey weights to adjust for unit nonresponse bias (e.g., Rothbaum et al. (2021)). AR have been used to mitigate known errors in survey data. For example, the Survey of Income and Program Participation (SIPP) replaces responses with AR when there are obvious errors to Supplemental Security Income (SSI) and Old-Age, Survivors and Disability Insurance (OASDI or Social Security) reporting driven by changes in the order and sequencing of health insurance coverage questions.[10] In the 2020 Census Nonresponse Followup (NRFU) operation, a mix of traditional fieldwork and AR enumeration was employed. AR were used to enumerate nonresponding housing units where AR were deemed to be of sufficient quality, after an enumerator made at least one attempt to obtain an in-person interview. Housing units with a predicted probability of being unoccupied over a certain threshold were classified as being vacant or delete (i.e., nonexistent or could not find).[11]

In this technical report, we describe a method using AR and statistical modeling to fill in citizenship where the American Community Survey (ACS) value is missing on a responding survey (item missing) and to construct household rosters and respective responses for the citizenship, age, race, and Hispanic origin of individuals within households that did not respond (unit missing). We use this method to create a Citizen Voting-Age Population (CVAP) special tabulation. We compare results from this enhanced method to those obtained using the ACS with the person weights employed in the past (hereafter called traditional weights), as well as tabulations that incorporated entropy balance person weights informed by auxiliary information in the processing of the 2020 data, described below.[12] Incorporating entropy balance weights in the processing of the 2020 data makes little difference to the citizen shares. The enhanced CVAP citizen share is 0.74 percentage points lower than that when incorporating entropy balance weights. The main contributor to this difference is the use of AR in nonresponding units rather than reweighting the responses to account for unit nonresponse. Replacing

[9] See Shin (2021).

[10] See Giefer et al. (2015).

[11] See Mulry et al. (2021) for a description of how AR were used in NRFU.

[12] We use the term “traditional weights” to refer to traditional weights applied to 2016-2020 and “entropy balance weights” to traditional weights applied to 2016-2019 and final weights incorporating entropy balance weights in the processing of the 2020 data. The latter version was used for the 2016-2020 5-year ACS and CVAP releases.

ACS citizenship edits and imputations with AR plays a more modest role. The group with the largest difference in citizen shares across methods is voting-age Hispanics, who have a 3.05 percentage point lower estimated citizen share with the enhanced CVAP method. The noncitizen coverage gap between AR and the ACS is greatest among those with unknown immigration status – people who may be particularly reluctant to participate in the ACS.

## *Background*

Since 2011, the U.S. Census Bureau has produced yearly special tabulations of the Citizen Voting-Age Population (CVAP) based on responses from the 5-year ACS. These tabulations were developed at the request of the U.S. Department of Justice to support its enforcement of Section 2 of the Voting Rights Act of 1965.

The emergence of the global COVID-19 pandemic in early 2020 caused major disruptions to data collection and other survey operations of the ACS. Instructions in a March 17, 2020, Office of Management and Budget (OMB) memo resulted in significant staffing reductions at the Census Bureau's National Processing Center (NPC) and telephone centers and required new interviewing procedures for both housing units and group quarters that emphasized public health and safety.

Table 1.1 Pandemic impact to 2020 ACS housing unit data collection

| Panel | Mailout Strategy | | | | | CAPI Operation[2] | | |
|---|---|---|---|---|---|---|---|---|
| | Initial Mailing | Reminder Letter | Questionnaire Package | Reminder Postcard | Final Reminder | Interviewing Month | Impact to Interviewing | CAPI letter sent |
| January | ✓ | ✓ | ✓ | ✓ | ✓ | March | In-person stopped on March 20 | |
| February | ✓ | ✓ | ✓ | ✓ | ✓ | April | Telephone only | |
| March | ✓ | ✓ | ! | ✗ | ✗ | May | Telephone only - increased workload | |
| April | ✗ | ✗ | ✗ | ✗ | ✗ | June | Telephone only - increased workload | |
| May | ✗ | ✗ | ✗ | ✗ | ✗ | July | In-person available - Some areas | ✓ From Regional Office |
| June | ✗ | ✗ | ✗ | ✗ | ✗ | August | In-person available - Some areas | |
| July | ✓ | ✗ | – | ✗ | ✗ | September | In-person available - All areas | |
| August | ✓ | ✗ | – | ✗ | ✗ | October | In-person available - All areas | |
| September | ✓ | ✗ | – | ✗ | ✗ | November | In-person available - Most areas | |
| October | ✓ | ✗ | ✓ | ✗ | ✓ | December | In-person available - Most areas | ✓ From NPC |
| November | ✓ | ✗ | ✓ | ✗ | ✓ | January '21 | In-person available - Most areas | |
| December | ✓ | ✗ | ✓ | ✗ | ✓ | February '21 | In-person available - All areas | |

✓ Mailed ! Part of workload mailed – Part of workload received questionnaire, remainder received reminder letter ✗ Was not mailed

Source: Shin (2021).

Table 1.1 summarizes the changes in data collection procedures for housing units.[13] The measures taken in response to the March OMB memo immediately affected the ACS mailout strategy. NPC was only able to mail out part of its March workload, and there were no mailings for the April, May, or June panels. The main ACS nonresponse follow-up operation, which is computer-assisted personal interviewing (CAPI), was also affected. Field representatives could only perform telephone-based interviewing for part of March and for the months of April through June. As a result of these restrictions, the 2020 housing unit response rate was 71.2 percent, which is much lower than the response rates in earlier years (94.7, 93.7, 92.0, and 86.0 percent in 2016, 2017, 2018, and 2019, respectively).

Data collection from group quarters was also affected. From mid-March through June, interviews were suspended completely. When interviews resumed, many nursing and correctional facilities kept entry restrictions in place to combat the pandemic, and many college and university dorms remained closed to students. Response rates for group quarters declined from 92 percent in 2019 and early 2020 to 55 percent in the months when interviews were suspended, then rose to about 70 percent for the remainder of the year.[14]

The Census Bureau attempted to remediate this additional nonresponse error by using a modified set of survey weights for the calculation of estimates based on the 2020 ACS. After a full demographic review of the weighted estimates, however, the data were deemed to not meet the Census Bureau's statistical quality standards due to the persistence of nonresponse bias. For example, the estimated noncitizen population in 2020 was 1.6 million lower than in 2019, while typical annual fluctuations are less than half a million. Many other demographic items showed similar effects. Because of these anomalies, the Census Bureau released only an experimental subset of standard 1-year ACS data products.

The project described in this report was commissioned when it became clear that the 1-year ACS data did not meet statistical quality standards even with modified weights designed to address additional nonresponse. The CVAP special tabulations use 5-year ACS data to support tabulations at lower levels of geography, but it was unclear at the beginning of this project whether the traditional 2016-2020 5-year ACS data would meet quality standards and be released as planned. Ultimately, the 2016-2020 5-year ACS using standard methods and a modified set of entropy balance weights for the 2020 ACS data met Census Bureau standards, and the 2020 CVAP special tabulation used those data.[15]

---

[13] See Shin (2021) for a detailed description of the normal ACS data collection procedures, the changes in 2020, and the effects of the COVID-19 pandemic on the estimates. Table 1.1 in this report is a replicate of Table 2.1 in Shin (2021).

[14] See Shin (2021).

[15] The entropy balance weighting method uses linked data from a variety of survey and administrative sources on both responding and nonresponding housing units, including demographics, household structure, income, employment, financing, and household characteristics. It aims to address unit nonresponse bias in not only citizenship, but also many other housing unit and demographic characteristics. See Rothbaum et al. (2021) for further details about entropy balance weights and U.S. Census Bureau (2022) about how they were incorporated into the 2016-2020 5-year ACS. One could consider developing an entropy balance weighting procedure that is

Survey nonresponse bias was present prior to the pandemic, and it will persist in the future.[16] It is thus informative for future planning to compare using AR to address item and unit nonresponse to traditional survey methods applying edits, imputations, and weight adjustments. Each method has potential sources of error or bias. The traditional survey method uses design-based estimates from self-response to the ACS. Item nonresponse for citizenship, race, and Hispanic Origin is handled via assignments (which use self-response data from the respondent or other household members to make logical inferences about missing responses) or allocation from hot-deck imputation (which generally involves borrowing values from geographically nearby self-responses that are judged to be demographically similar). Unit nonresponse is mainly handled by adjusting the survey weights for the responding units to resemble the whole population more closely.

As-reported citizenship values potentially suffer from social desirability bias, where respondents believe it is less risky to report noncitizens as citizens due to social pressures. Item nonresponse may be non-random, where noncitizens may be more likely to skip the question due to language and access barriers or fear and distrust among immigrant communities.[17] The traditional method of imputing missing citizenship values by replacing them with self-responses from nearby, demographically similar individuals may inflate the count of citizens; it is likely that individuals who leave the citizenship question blank are systematically different from those who give a response not requiring an edit or imputation, even after accounting for differing demographic characteristics.

Households with noncitizens could also be less likely to respond to the survey at all, leading to unit nonresponse bias that is difficult to correct through weighting adjustments to responses, as has been traditionally done. Respondents could also omit vulnerable noncitizens from the household roster, making weight adjustments to rosters suffering from such errors less effective. All these potential error sources suggest that the traditional method of producing the CVAP leads to an upward bias in the counts of citizens, as also suggested by previous research (Abowd et al. 2020, Brown et al. 2019b, and Jensen et al. 2015). Abowd et al. find that the citizen share of the population using ACS allocation values is significantly higher than that in high-quality AR. Jensen et al. (2015) suggest that the ACS may underestimate the number of foreign-born Hispanics, even after applying population controls intended to mitigate nonresponse bias, because the population controls do not reflect citizenship or country of birth. Van Hook et al. (2014) estimate undercoverage of Mexican-born immigrants in the 2000 Census and ACS, especially among those who likely had undocumented immigration status. Brown et al. (2019b) demonstrated that households containing noncitizens were less likely to respond to the ACS.

The enhanced CVAP method described in this report also has potential sources of error. Record linkage errors could lead to assignment of AR citizenship to a different person in the ACS who may not have the same citizenship status. The direction of the bias from such errors is hard to predict. Certain

specifically tailored to citizenship and addresses both unit nonresponse bias and omissions from the ACS household rosters, building on Zaslavsky (1988). We leave this for future research.

[16] For more information, see, for example, page 38 of: https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf.

[17] See Tourangeau and Yan (2007) for discussion of social desirability bias regarding responses to sensitive questions, and Evans et al. (2019) and McGeeney et al. (2019) for survey evidence of concern by foreign-born respondents about how the Census Bureau may use their response data.

populations, such as undocumented immigrants, are not as well-represented in AR, and it may be more difficult to successfully link their ACS data to AR using production record linkage systems.

At the outset of this project, we decided that any reported citizenship value in the ACS that did not violate the traditional edit rules would be accepted as-is and not be overruled by information from AR.[18] To the extent that the traditional method is subject to social desirability bias, the enhanced method will reflect this bias as well. Similarly, our enhanced CVAP method does not alter the household rosters for responding units. This means that the enhanced CVAP method does not correct for omissions of noncitizen household members in responding units. These sources of error also tend to create an upward bias in citizen estimates. AR may not contain a person's complete set of addresses, and the ones the person has generally do not show exactly where they reside each month. This leads to uncertainty about whether the person resides in a particular ACS-sampled housing unit or not. This uncertainty is likely to be a greater issue for noncitizens, since they are ineligible for many of the government programs that generate our AR files, so estimates for groups containing more noncitizens may be less precise. There is also uncertainty about whether an AR person is a U.S. resident in the ACS tabulation month. The probability of an AR person residing outside the U.S. in that month is likely to be higher for noncitizens (e.g., an international student who studied at a U.S. university for a period of time, then returned to their home country).

In Section 2, we describe the data and methods used to create the enhanced CVAP tabulation, including record linkage; assembling the AR people to include in the dataset; creating their age, race/ethnicity, citizenship, and person-place probabilities; and tabulating the data. We compare the enhanced CVAP tabulation to ones using traditional and entropy balance weights in Section 3. Section 4 concludes.

# 2 Data and Methods

## *2.1 Record Linkage*

Record linkage plays an essential role in bringing together multiple sources of information about individuals in this project. Here we describe the Census Bureau production process for assigning unique person identifiers to data records that are used in this report.

The Census Bureau's Person Identification Validation System (PVS) assigns Protected Identification Keys (PIKs) to individuals.[19] PIKs are anonymous, unique person identifiers that are temporally invariant just like Social Security numbers (SSNs). SSNs are replaced by PIKs in files that initially contain SSNs when received by the Census Bureau, since access to files containing SSNs is limited to a small staff that specializes in maintaining the record linkage system. This process facilitates linking person records across files while protecting individual PII from distribution within the Census Bureau.

[18] This decision followed senior Census Bureau leadership guidance to research enhancements that were strongly consistent with current methodology, which does not generally overwrite self-responses unless they fail edit rules.

[19] For details see Wagner and Layne (2014).

The PVS uses probabilistic record linkage (Fellegi and Sunter, 1969) to link data from an incoming file (e.g., a survey or AR file) to reference files containing data on SSN applications from the NUMIDENT enhanced with address data obtained from other federal administrative records. ITINs and the names and addresses associated with them in federal administrative records are also included in the PVS process.[20] ITINs are issued to people who need to make tax filings to the Internal Revenue Service, but who are not eligible to have an SSN. According to Gee et al. (2017), at least 50 percent of undocumented immigrant households file income tax returns using ITINs.

People who do not have an SSN or ITIN are not in the PVS reference files and thus would only receive a PIK in error. Since the need for U.S. residents to have SSNs is nearly universal, almost all U.S. citizens should have them.[21] In a study comparing AR to the entire 2020 Census, Brown et al. (2023) show that the U.S. residents without SSNs or ITINs are predominantly noncitizens, and the citizen share is lower when incorporating them in the calculations. That suggests that not including people without SSNs or ITINs in the data used in this report creates an upward bias on the citizen share.

## *2.2 Compiling Administrative Data on People Linked to ACS Housing Units*

To construct administrative CVAP data using the ACS as the frame, we compiled AR data for the responding and nonresponding housing units (HUs) in the initial 2016-2020 ACS HU sample.[22,23] Specifically, we constructed a linked survey-administrative dataset where we populated each ACS HU with people observed in AR living at the HU around the time when the HU completed or was expected to complete its ACS interview. The month when an HU completed or was expected to complete its ACS interview is called its "tabulation month." HUs that complete their ACS interview are included in the data for their respective tabulation month.

Throughout this section we will refer to variables obtained from ACS datasets as "ACS" variables and to variables obtained from administrative datasets as "administrative" variables.

### *Step 1: From the initial 2016-2020 ACS HU sample, create a roster of unique HU-year pairs*

We began with the initial 2016-2020 ACS HU sample, which included all responding and non-responding ACS HUs with a tabulation month between January 2016 and December 2020. We retained the ACS year, the tabulation month, and the identifier (called a MAFID) assigned to the HU

[20] The Census Bureau does not have access to the ITIN application data. The range of nine-digit numbers in the Taxpayer Identification Number field used for ITINs is public information. In subsequent sections of this report, we refer to people with numbers in that range as having ITINs.

[21] See Puckett (2009).

[22] We did not construct administrative data for people in group quarters, other than to replace ACS citizenship that was not as-reported with administrative record citizenship, if available, or modeled citizenship probabilities otherwise.

[23] Responding housing units include those classified as occupied, vacant (except those classified as vacant through an occupancy prediction model), temporarily occupied (vacant in the final tabulation), or delete. Nonresponding housing units are those classified as vacant through an occupancy prediction model, non-interviews (e.g., refusals), and those that both did not self-respond and were not selected for Computer-Assisted Personal Interview (CAPI) fieldwork follow-up.

from the Census Bureau's Master Address File (MAF). Records in the resulting dataset were uniquely identified by MAFID and year.[24] We will refer to this dataset as the "ACS MAFID-year roster."

### *Step 2: Identify sources of administrative data with records of uniquely identified individuals paired with uniquely identified addresses (PIK-MAFID pairs)*

Next, we identified sources of administrative data that placed people at addresses between 2015 and 2021. Specifically, we identified all sources of administrative data held by the Census Bureau that met the following criteria: (1) the data placed people at addresses, (2) PIKs were assigned to people, and MAFIDs were assigned to addresses, (3) the data contained enough information to assign reference years to observations of PIK-MAFID pairs, and (5) the data contained observations of PIK-MAFID pairs from at least one year between 2015 and 2021. We included in our linked survey-administrative dataset any administrative data that met criteria (1)-(5) and were approved for use on this project.[25] Table 2.1 provides a description of each source of administrative data and shows the years covered by each source.

### *Step 3: Assign dates to administrative records of PIK-MAFID pairs*

Most sources of administrative data provided enough information to assign reference months within reference years, and many provided enough information to assign reference days within reference months. We assigned full dates to observations of PIK-MAFID pairs whenever possible. If a source provided enough information to assign reference years and months but not days, we assigned full dates using the years and months from the source and set the day of every observation equal to 15, the midpoint of most months.

### *Step 4: Create the administrative PIK roster (a roster of unique ACS MAFID, ACS year, administrative PIK triples)*

We linked each MAFID-year pair from the ACS MAFID-year roster to every AR of a PIK-MAFID pair where (1) the administrative MAFID linked to the ACS MAFID and (2) the administrative month was within a 25-month window centered on the ACS tabulation month (the tabulation month plus 12 months on either side) or within a 3-year window centered on the ACS year. When processing AR from sources that provided both reference months and years for observations of PIK-MAFID pairs, we used the 25-month window; when processing AR from sources that provided reference years only, we used the 3-year window. These windows were chosen to balance the competing goals of capturing AR PIKs of as many of the people who were present at an ACS HU in its tabulation month as possible and excluding AR PIKs of people who were not present in the ACS HU.

---

[24] Records were uniquely identified by MAFID and year together rather than by MAFID alone because a small number of housing units (HUs) with a tabulation month in 2016 had a second tabulation month in 2020. Consistent with ACS sampling procedures, no HU was sampled more than once between 2016 and 2020. However, a HU sampled for the ACS in a given month and year may have a tabulation month up to two months later than its sample month. For this reason, some HUs sampled in November or December 2015 had a tabulation month in January or February 2016. Some of these HUs were sampled again in 2020 and had a second tabulation month in 2020.

[25] Work on this project took place under the Census Bureau's Data Management System (DMS) No. 7505696.

When doing the linking, we sometimes encountered the same PIK-MAFID pair on records from multiple AR sources or on multiple records from the same source. We unduplicated the records in the resulting linked dataset by ACS MAFID, ACS year, and administrative PIK to produce a dataset with records uniquely identified by these variables. We will refer to this dataset as the "administrative PIK roster." Each record on the administrative PIK roster also included the tabulation month of the given ACS MAFID in the given ACS year.

### *Step 5: Apply birth, death, and incarceration filters to records on the administrative PIK roster*

We dropped from the administrative PIK roster any ACS MAFID, ACS year, administrative PIK triple that met any of the following criteria: (1) based on the death date associated with the PIK in the NUMIDENT, the person died before the tabulation month of their ACS MAFID, (2) based on the birth date associated with the PIK in the NUMIDENT, the person was born after the tabulation month of their ACS MAFID, or (3) based on the data the Census Bureau received from the Bureau of Prisons, the person was incarcerated during the tabulation month of their ACS MAFID.

### *Step 6: Use the Census Household Composition Key (CHCK) to create the CHCK PIK roster*

The CHCK is a crosswalk that links the PIKs of children ages 0 to 19 each year to the PIKs of their mother and father. The Census Bureau creates CHCK using data from SSA and previous decennial censuses. We used the 2016-2020 versions of the CHCK to identify children whose primary parent appeared on the AR PIK roster in a given ACS year but who did not themselves appear on the roster in that year. Because children are more likely to live with their mother than their father when the parents live separately, we defined a child's primary parent as the mother if the mother's PIK was available on the child's CHCK record, and the father if the mother's PIK was not available. We excluded children who turned 18 before the tabulation month of their primary parent's ACS MAFID. Records in the resulting dataset were unique by ACS MAFID, ACS year, CHCK parent PIK, and CHCK child PIK. We refer to this dataset as the "CHCK PIK roster."

### *Step 7: Create the as-reported administrative demographics dataset*

We assigned age to people (PIKs) on the administrative PIK roster and to children (child PIKs) on the CHCK PIK roster using information from the 2020 fourth quarter NUMIDENT file. We assigned each person an age as of the last day of the tabulation month of their ACS housing unit (MAFID) using their date of birth from the NUMIDENT. If the computed age value exceeded 114, the age of the oldest person alive in the United States in 2020,[26] we assumed it was inaccurate and overwrote it with a missing value. We also assigned sex from the NUMIDENT.[27]

We assigned race and ethnicity using information from the Census Best Race File, which uses information from AR, household survey data, decennial census data, and third-party data to assign a

[26] See https://en.wikipedia.org/wiki/List_of_American_supercentenarians.

[27] Though sex is not a variable used in the CVAP tabulations, it is an input to the modeling of the other demographic characteristics.

race category and an ethnicity category to people who have been assigned a PIK.[28] The race and ethnicity categories on the Best Race File differ somewhat from the categories used in the standard CVAP tabulations. We recoded the Best Race categories to match the ACS-CVAP categories: (1) non-Hispanic American Indian or Alaska Native (AIAN) alone, (2) non-Hispanic Asian alone, (3) non-Hispanic Black or African American alone, (4) non-Hispanic Native Hawaiian or other Pacific Islander (NHPI) alone, (5) non-Hispanic White alone, (6) non-Hispanic AIAN and White, (7) non-Hispanic Asian and White, (8) non-Hispanic Black or African American and White, (9) non-Hispanic AIAN and Black, (10) non-Hispanic remainder or two or more racial categories, and (11) Hispanic. We refer to the resulting dataset as the "as-reported administrative demographics dataset."

When we could not assign a demographic characteristic to a PIK on the administrative PIK roster or a child PIK on the CHCK PIK roster, we used other characteristics of the PIK and its ACS MAFID to impute the demographic characteristic. Among the characteristics used in the imputation process were a set of variables capturing aggregate information from the 2010 Census about the geographic areas (state, county, tract, and block) where the ACS MAFID was located. To facilitate the creation of these aggregates, we added to the as-reported administrative demographics dataset the state, county, tract, and block of each ACS MAFID. We used geographic information from the raw pre-disclosure avoidance ACS data files where available; otherwise, we used the 2010 tabulation geographies from a July 2020 MAF extract for the ACS.

### *Step 8: Create the administrative PIK-MAFID roster (a roster of unique ACS MAFID, ACS year, administrative PIK, administrative MAFID quadruples)*

We linked each ACS MAFID, ACS year, administrative PIK triple from the administrative PIK roster to every AR PIK-MAFID pair where (1) the PIK on the administrative PIK roster linked to the PIK on the AR PIK-MAFID pair and, (2) the administrative month was within a 25-month window around the ACS tabulation month or within a 3-year window around the ACS year. (See step 4 for an explanation of these time windows.) Here the AR MAFIDs were not restricted to being ACS MAFIDs.

When doing the linking, we sometimes encountered the same PIK-MAFID pair on records from multiple administrative data sources or on multiple records from the same source. We unduplicated the records in the resulting linked dataset by ACS MAFID, ACS year, administrative PIK, and administrative MAFID to produce a dataset with records uniquely identified by these variables. We refer to this dataset as the "administrative PIK-MAFID roster."

Each record on the administrative PIK-MAFID roster also included (1) the tabulation month of the given ACS MAFID in the given ACS year, (2) the latest date up to and including the ACS tabulation month when the administrative PIK was observed at the administrative MAFID, (3) the earliest date after the ACS tabulation month when the administrative PIK was observed at the administrative MAFID, (4) the latest year before or coincident with the ACS year when the administrative PIK was observed at the administrative MAFID, (5) the earliest year after the ACS year when the administrative PIK was

[28] For more information about the process for assigning race and ethnicity in the Census Best Race File, see Ennis et al. (2018).

observed at the administrative MAFID,[29] and (6) an indicator variable for each AR source that equaled one if the administrative PIK was observed at the administrative MAFID within the focal window in the data from that source.[30]

We used predictive models to predict the probability that each PIK we observed in the administrative data at an ACS MAFID in a window around the tabulation month of the MAFID lived at that MAFID during the tabulation month. To facilitate this person-place modeling, we added to the administrative PIK-MAFID roster some additional variables from several administrative data sources.[31] (See Section 2.6.)

### *Step 9: Create the person-place modeling (PPM) input dataset*

We modeled the probability that each PIK on the administrative PIK roster was at its ACS MAFID in its ACS tabulation month using a dataset derived from the PIK-MAFID roster. First, we collapsed the source indicator variables, date variables, and year variables from the state SNAP, WIC, and TANF datasets to treat these sources as a single source. Next, we collapsed the source indicator variables, date variables, and year variables from the state driver's license files. Finally, we added two indicator variables: (1) a variable that equaled one if the ACS MAFID on the record represented a responding ACS HU from the ACS year on the record, and (2) a variable that equaled one if the administrative PIK-MAFID pair on a record linked to an ACS PIK-MAFID pair from the ACS year on the record. We refer to the resulting dataset as the "person-place modeling (PPM) input dataset."

## *2.3 Age, Sex, and Race/Ethnicity Modeling*

In this section, we describe the modeling processes used to determine age, sex, race, and ethnicity for observations in an AR-augmented 2016-2020 ACS housing unit frame. While the use of a model for determining demographic characteristics in this population was only for people in nonresponding ACS HUs when constructing enhanced CVAP estimates, this procedure provided a predicted probability for every individual observation in the frame. This demographic modeling process involved separate models for sex, age, and race and ethnicity. Each model was fit using three sets of covariates: address- and geography-level means from AR data, geography-level means from the 2010 Census, and individual-level probabilities based on first and/or last names. Each of these models was trained on a 10 percent sample of the data frame. The resulting model coefficients were then used to calculate the

[29] We included both year and date variables rather than date variables alone because every record in the administrative PIK-MAFID roster had at least one non-missing year value, while only some records had no non-missing date values.

[30] Some sources of administrative data listed in Table 2.1 provided multiple types of addresses, and in some cases, we created separate source indicators for the different types of addresses. Specifically, we created separate source indicators for addresses in single- and multiunit properties in the Federal Housing Administration (FHA) data; for addresses "moved from" and "moved to" in National Change of Address (NCOA) data; and for facility addresses for people in custody, facility addresses for people received into custody, alternate addresses for people in custody, and alternate addresses for people received into custody in the U.S. Marshals (USMS) data.

[31] The additional variables included a variable distinguishing pre- from post-prison addresses in the BOP data, a set of variables capturing information about the housing unit from the HUD data, a set of variables capturing information about the tax unit from the IRS 1040 data, a set of variables capturing information about the move from the NCOA data, and a set of variables capturing information about the household from the VSGI data. For the full name and description of the data source identified by each abbreviation, see Table 2.1.

predicted probability of each demographic categorical outcome for each observation in the frame. These predicted probabilities can be used to tabulate the expected population with said characteristics for any group of observations or for randomly drawing imputes.

These demographic models used as-reported administrative demographics (described in Section 2.2) as the dependent variable. *Age* was a binary variable indicating 18 and over or under 18. *Sex* was a binary variable indicating male or female. Race and ethnicity were consolidated into a single variable, *CVAP_race*, containing the 11 CVAP categories of race and ethnicity.

The model used three sets of covariates, including individual, address-level, and geography-level means of AR-based characteristics, geography-level means of characteristics from 2010 Census data, and name-based probabilities from SSA names data. In each model, the individual AR age, sex, and race/ethnicity variables that were not the dependent variable were included as covariates. For instance, for the age model, sex and race/ethnicity categorical variables were covariates. The address-level and geography-level aggregates of AR characteristics included the proportion male, the proportion 18 and over, and the proportion in each race/ethnicity category (excluding the reference category of non-Hispanic White alone) at the address, block group, tract, county, and state level. If a geography or proportion was missing, the value was filled with the next higher level of aggregation. For instance, if the tract level proportion of Hispanic individuals was missing, the covariate was replaced by the county-level proportion Hispanic.

The model also used geographic variables based on 2010 Census responses.[32] The proportion male, the proportion 18 and over, and the proportion of each race and ethnicity category (excluding the reference category of non-Hispanic White alone) were calculated using the 2010 Census Edited File. These covariates were constructed at the state, county, tract, and block level for every observation. If a geography or proportion was missing for an observation, the value was filled with the non-missing value of the next higher level of aggregation.

Lastly, the model used predicted probabilities which were output from a set of name-based models of demographic characteristics as an additional covariate. For age, the probabilities that the individual's age were in the following bins (0-5, 6-14, 15-17, 18-24, 25-34, 45-54, 55-64, 65-74, 75+) were calculated from the SSA names data. The data were collapsed by first name and age bins to derive the probability of being a member of a given age group given one's first name. The predicted probabilities of being in each age bin were the set of variables *prob_age_n* where *n = (0, 6, 15, 18, 25, 45, 55, 65, 75).*

For sex, first and middle names were split on any separators (space or hyphen) and assigned to the new variables *first_name_1*, *first_name_2*, *first_name_3*, *middle_name_1*, *middle_name_2*, and *middle_name_3*. Using the gender command in the *gender* package in R,[33] we assigned a male probability to each name variable and kept the male probability from the first name (the original first name variable prior to splitting) if non-missing, then filled the remainder by middle name (the original middle name variable prior to splitting), then by *first_name_1*, etc., through *middle_name_3*. The predicted probability of being male from this model was saved as the variable *prob_male*.

[32] We used 2010 Census data since the ACS data from this timeframe also uses 2010 geography codes.

[33] For information on the gender package, see Blevins and Mullen (2015). See also the package URL https://github.com/lmullen/gender. For more information on R, please refer to R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

For race and ethnicity, the NUMIDENT data were collapsed by surname and race and ethnicity for each of seven race categories and Hispanic/non-Hispanic dummies. The seven race categories were (1) White alone, (2) Black alone, (3) Asian alone, (4) American Indian or Alaska Native (AIAN) alone, (5) some other race (SOR) alone, (6) Native Hawaiian or other Pacific Islander (NHPI) alone, and (7) two or more races. Each observation was attached with the probability of a person with that surname being in each race category as well as the probability of being Hispanic. The predicted probability of being in each race/ethnicity category from this model was saved as the variable *probbr_n.*[34]

Each model was a logistic regression with the following specifications:

- *Age* was a binary dependent variable which indicated the person was 18 or older. We ran the logistic regression specification

$$logit(Age_i) = ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta,$$

where $x_i$ was a vector of individual covariates which included an intercept, indicators for *CVAP_race*; *Sex*; name-based probabilities *prob_age_n*, where *n = (0, 6, 15, 18, 25, 45, 55, 65, 75)*; state, county, tract, block group, and MAFID-level means of the *CVAP_race* indicators, *Age*, and *Sex* from AR; and state-, county-, tract-, and block-level means of the *CVAP_race* indicators, *Age*, and *Sex* from the 2010 Census. The probability the person was 18 or older was

$$\pi_i = P(Age_i = 1|x_i) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}.$$

- Similarly, *Sex* was a binary dependent variable that indicated the person was male. We ran the logistic regression specification

$$logit(Sex_i) = ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta,$$

where $x_i$ was a vector of individual covariates which included an intercept; indicators for *CVAP_race*; *Age*; name-based probability *prob_male*; state-, county-, tract-, block group-, and MAFID-level means of the *CVAP_race* indicators, *Age*, and *Sex* from AR; and state-, county-, tract-, and block-level means of the *CVAP_race* indicators, *Age*, and *Sex* from the 2010 Census. The probability the person was male was

$$\pi_i = P(Sex_i = 1|x_i) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}.$$

[34] These predicted probabilities were estimated for a different project. We reused them here due to time constraints. Ideally, the CVAP race categories should be used.

- For *race and ethnicity*, we ran a similar multinomial logistic regression specification where the dependent variable, *CVAP_race*, had the 11 categories of race and ethnicity enumerated above. Specifically, we estimated

$$logit\left(CVAP_{race_{im}}\right) = ln\left(\frac{P\left(CVAP_{race_i} = m\right)}{P\left(CVAP_{race_i} = 1\right)}\right) = x_i'\beta_m, \quad m = 2, \ldots, 11,$$

where $m$ was each of the 11 categories of *CVAP_*race, excluding the reference category *CVAP_race* = 1 (non-Hispanic white alone). $x_i$ was a vector of individual covariates which included an intercept; *Sex*; *Age*; name-based probabilities *probbr_n*; state-, county-, tract-, block group-, and MAFID-level means of the *CVAP_race* indicators, *Age*, and *Sex* from AR; and state-, county-, tract-, and block-level means of the *CVAP_race* indicators, *Age*, and *Sex* from the 2010 Census. The probability that *CVAP_race* = *m* was

$$P\left(CVAP_{race_i} = m \middle| x_i\right) = \frac{e^{x_i'\beta_m}}{\sum_{j=1}^{11} e^{x_i'\beta_j}}, \quad m = 1, \ldots, 11.$$

These models were trained on a 10 percent sample of the data with non-missing values of the dependent variable. A predicted probability was calculated for each observation in the frame from the coefficients of the fitted model, as described above. For the race/ethnicity model, the output was a predicted probability for each of 11 categories of race and ethnicity.

Note that the models we used assumed that observations for which the dependent variable was missing did not differ in terms of their relationship with the covariates from the observations for which the dependent variable was populated (missing conditionally at random). This assumption may be violated, for instance, if demographic characteristics are correlated with lack of coverage in administrative sources. Given the high level of coverage and quality of SSA data and the focus of this exercise on individuals who can be linked by PIK, the extent of missingness in the dependent variables was small.

## *2.4 Citizenship Business Rules*

This section describes all the citizenship data sources and how they were combined to construct the AR citizenship variable for people in ACS responses in both housing units and group quarters, as well as AR people with addresses in the ACS housing unit frame.[35]

The NUMIDENT is a record of individual applications for SSNs and subsequent updates for those people. All U.S. citizens are eligible to have an SSN, which is required to legally work, to receive Social Security benefits, and to receive other federal-government-administered social services. The NUMIDENT provides the most comprehensive coverage of U.S.-born people. NUMIDENT coverage of noncitizens is less complete, as generally only those authorized by the Department of Homeland Security to work are eligible for an SSN.

[35] For a more detailed description, see Abowd et al. (2020).

In May 1981, SSA began recording citizenship information in the NUMIDENT. For people who applied for an SSN prior to that date and who have not subsequently updated their SSN record, the only available citizenship-related information is place of birth and an associated foreign country of birth for those not born in the United States.

SSA is not automatically notified when previously noncitizen SSN holders become naturalized citizens, so naturalizations may be captured with a delay or not at all. Consistent with this, Brown et al. (2019a) show that NUMIDENT citizenship changes usually occur sometime after the year of naturalization.

We addressed this SSA data gap by supplementing it with the sources used to document a naturalization, namely, U.S. Citizenship and Immigration Services (USCIS) naturalization certificates and U.S. passports. People naturalized as adults must obtain a naturalization certificate. Children under age 18 eligible for naturalization when their parents become naturalized (derived citizenship) can document their naturalization either through a naturalization certificate or a U.S. passport.

The passport data we used, provided by the Department of State, covered all U.S. passports issued between 1978 and April 1, 2020. Though passports are considered the most definitive evidence of citizenship,[36] and all citizens are eligible for a passport, many do not have one. The data coverage skews towards those who travel outside the U.S. Passports complement the NUMIDENT and USCIS coverage of citizens. The passport data not only helped fill the gap for people naturalized through derived citizenship, but also for some foreign-born people in the NUMIDENT who applied for an SSN prior to May 1981. Note that naturalized citizens are more likely to hold passports than U.S.-born citizens, e.g., so that they can visit family members outside the U.S.

The USCIS data included all available electronic records through April 1, 2020, for approved applications for naturalization and lawful permanent residence, as well as asylum and refugee data.

Not all U.S. residents are eligible for SSNs, since one must be a U.S. citizen or have a valid work permit. To gain partial coverage of the non-SSN population, we used the ITIN indicators discussed in Section 2.1. Once ITIN-holders become eligible for an SSN, they are no longer allowed to use the ITIN. Due to that fact, and that all citizens are eligible to have an SSN, ITIN-holders must be noncitizens during the time they use the ITIN.

We considered the above data sources as primary sources because they require documented proof of citizenship status. To fill in coverage gaps, we supplemented them with data where citizenship verification may be either less stringent or nonexistent. These secondary sources included data from the Federal Bureau of Prisons (BOP), U.S. Marshals Service (USMS), state driver's licenses, state public assistance programs, and Census Bureau household surveys.

The BOP data contained records of inmates in federal prisons from 1980 to April 1, 2020. The data included the date the inmate was last in federal prison.

The USMS data included records for federal prisoners in the custody of USMS. The data came from the USMS Prisoner Processing and Population Management/Prisoner Tracking System, and they covered 2010 to April 1, 2020.

[36] See https://www.uscis.gov/us-citizenship/proof-us-citizenship-and-identification-when-applying-a-job.

Driver's license and state identification card citizenship data covered Alabama, Nebraska, South Carolina, and South Dakota.[37]

Supplemental Nutrition Assistance Program (SNAP) and Temporary Assistance for Needy Families (TANF) files for eight states contained citizenship information that we used here. We also included program denial due to citizenship status information from five states.

We included as-reported citizenship responses in Census Bureau surveys conducted prior to the ACS tabulation month. The surveys included not only the ACS, but also the Current Population Survey (CPS), the American Housing Survey (AHS), and the Survey of Income and Program Participation (SIPP). Survey responses were not verified, and they were current as of the time of the interview, which in some cases took place many years ago.

The rules we used to construct a citizenship variable from these sources were as follows. A person was classified as a citizen if they had information suggesting they were a citizen in the NUMIDENT (either a U.S. citizen or missing citizenship but U.S.-born), a U.S. passport, and/or a USCIS naturalization certificate. People without evidence of being a citizen in the above sources and who had a noncitizen value in at least one of the primary or secondary noncitizen sources were classified as noncitizens. People with no information suggesting citizenship status in either of the above groups and who had a citizen value from a secondary citizen source were classified as citizens.[38] The balance of the citizenship responses were modeled.

## *2.5 Citizenship Modeling*

In developing alternative methods for estimating citizenship for the CVAP tabulations, we took two steps to modeling an individual's citizenship status when it was not self-reported.[39] The first step was to take AR citizenship (explained in Section 2.4) for those who had it, and then use predicted probabilities from logistic regression models for the remainder of the missing values. We trained the models on enhanced ACS data from 2015-2019.[40] In this section, we provide our motivation for the model design, describe the model estimation in detail, and provide descriptive statistics from the models.

People without AR citizenship can be divided into five groups.[41] The first group is people with PIKs but no AR citizenship (NAR-PIK). Most of these are foreign-born in the NUMIDENT, but their citizenship variable is missing from the record. It is missing because they applied for an SSN before May 1981, and

[37] We also use driver's license data from Iowa when constructing AR PIK-MAFID data. The Iowa data do not contain citizenship.

[38] Abowd et al. (2020) conducted analysis supporting this set of rules. They used a few secondary sources not available for this project, but the additional coverage those sources provided in that study is only about 0.1 percent.

[39] We produced modeled citizenship probabilities for all people in ACS housing units and those in group quarters, as well as AR people with an address in the initial ACS housing unit frame. In the enhanced CVAP method, we did not use the modeled probabilities for people with as-reported ACS citizenship or the AR citizenship variable described in Section 2.4.

[40] The enhanced ACS data include additional PIK-MAFID records described in the previous section.

[41] Hereafter we refer to the AR citizenship group and these five no-AR citizenship (NAR) groups as AR/PVS groups.

they have not updated their information with SSA. These people are generally long-term U.S. residents. They may have obtained a U.S. passport prior to 1978 (the first year of available passport data) and not renewed it since or could have a USCIS naturalization certificate issued many years ago, before the naturalization data coverage was complete.

The second group is people without PIKs who were sent to PVS search but failed to link to a valid PIK (NAR-SS). A record could fail PVS search due to discrepancies in how the PII is reported in the ACS versus other survey or administrative data, use of a different address, having a common name, or because the person is genuinely absent from the NUMIDENT and other PVS reference files. People absent from the reference files are highly likely to be noncitizens, since all citizens are eligible for an SSN. Regardless, both citizens and noncitizens could have discrepant PII in the survey data.

The third group is people without PIKs who were not sent to PVS search due to insufficient PII (NAR-NSS). Insufficient PII could occur if the respondent has confidentiality concerns. For example, they could be worried that the data may be used for individually targeted law enforcement. PII-deficient cases also often arise in the census when the respondent is a proxy, and the proxy does not know or will not report the neighbor's PII. These cases appear in household surveys where one respondent is responsible for reporting information about all household members, and the respondent may not know detailed information about a member of the household or could choose not to report them. There are also census and household survey count imputations and substitutions, which are person records that lacked any PII because all characteristics were imputed.

The fourth group is people who have PIKs, do not have AR citizenship, have AR listing an ACS address, but who are not in the ACS response data (NAR-NACS).

The fifth group is ACS people without PIKs who were not sent to PVS search due to confidentiality reasons (NAR-NPVS). This group is mostly group quarters records.

### *ACS Citizenship Logistic Regressions*

We estimated a logistic regression model on ACS data for each group mentioned above and used the estimated model to predict citizenship for the people in the same group in the 2016-2020 ACS (including housing units and group quarters), as well as AR people with addresses in the initial 2016-2020 ACS housing unit sample. In this exercise, we estimated models using enhanced ACS data for the years 2015-2019. The predictions were applied to the enhanced 2016-2020 ACS data.[42]

The dependent variable in each logistic regression model was the binary (citizen versus noncitizen) as-reported ACS citizenship; that is, for all missing citizenship models, we predicted the ACS response, not the AR value. For the first three groups (NAR-PIK, NAR-SS, NAR-NSS), we estimated models using the population in those groups in the training sample (2015-2019 ACS). For the NAR-NPVS group, we

[42] We chose not to use 2020 data to train the models due to the documented problems with citizenship responses in the 2020 ACS. We considered using a single year of ACS data to train the models (e.g., 2015 for the 2016-2020 ACS sample). This would have the advantage of not having overlapping training and prediction samples. However, this would come with the disadvantage of training models with data from periods farther removed to the reference date for the statistics (especially 2020). This could be important if associations between citizenship and other observable characteristics change rapidly. Another disadvantage of having only one year in the training sample is that the smaller sample size generates difficulties in the estimation of richer models.

estimated the model using group quarters records in the training sample. Lastly, for the NAR-NACS group, we estimated the model using the NAR-PIK sample. We did so because the NAR-NACS group was not in the ACS response data and therefore did not have the as-reported ACS citizenship needed to train the model.

For four of the five models (NAR-PIK, NAR-SS, NAR-NSS, and NAR-NPVS), we employed similar sets of explanatory variables. These included age groups (18-29, 30-49, and 50-plus); a female indicator; indicators for the CVAP race/ethnicity categories other than non-Hispanic Asian and Hispanic, Hispanic origin subgroups (Mexican, Puerto Rican, Cuban, Central American, Latin American, and other Hispanic), and Asian subgroups (Asian Indian alone, Chinese alone, Filipino alone, Japanese or Okinawan or Iwo Jiman alone, Korean alone, Vietnamese alone, or other non-Hispanic Asian); an indicator for whether the home is owned with or without a mortgage versus rented or occupied without rent; household size group indicators (1, 2, 3, 4, or 5 or more); six relationship to the householder categories (householder, spouse/partner, child, all other relatives, unrelated individuals, and group quarters); the shares of other household members by citizenship/PVS category (AR citizen, AR noncitizen, NAR-PIK/SS, combining NAR-PIK and NAR-SS, or NAR-NSS); the citizenship category of the householder interacted with being a relative of the householder and a non-relative of the householder; indicators for whether the language of the survey data was English, Spanish, another language, or missing (the missing category was included in the NAR-PIK and NAR-NSS regressions, while it was combined with English in the NAR-SS regression); the county ACS citizen share among those who were in the regression sample's AR/PVS group; and state indicators. The NAR-SS and NAR-NSS regressions included age-state, tenure-state, age-tenure, and relationship category-tenure interactions.

The NAR-SS model was the most extensive due to sample size and data availability. It included state indicators, age group indicators (18-29, 30-49, and 50+), an indicator for whether the home was owned with or without a mortgage versus rented or occupied without rent, household size group indicators (1, 2, 3, 4, or 5 or more), indicators for the full set of relationship categories (householder, spouse/partner, child, all other relatives, unrelated individuals, and unknown relationship), the shares of other household members in the NAR-SS and NAR-NPVS groups, indicators for the form language used (English, Spanish, another language, or missing), an indicator for being female, indicators for the CVAP race/ethnicity categories other than non-Hispanic Asian and Hispanic, Hispanic origin subgroups (Mexican, Puerto Rican, Cuban, Central American, Latin American, and other Hispanic), and Asian subgroups (Asian Indian alone, Chinese alone, Filipino alone, Japanese or Okinawan or Iwo Jiman alone, Korean alone, Vietnamese alone, or other non-Hispanic Asian), an indicator for living in group quarters, an indicator for records without a PIK that had equally probable links to multiple people in the reference files,[43] and interactions between state and age group, state and home ownership, home ownership and relation type, and the relative indicator and the citizenship category of the household.

The NAR-NSS model was identical to the NAR-SS model except for using Census Division instead of state and removing the group quarters indicator, the indicator for records that received multiple links to the reference files, and all interactions. The NAR-NPVS model was identical to the NAR-SS model except for removing the indicator for records that were not assigned a PIK due to having equally

[43] This indicator corresponds to records not receiving a PIK because the record from the source file links to two or more different PIKs in the reference files at the same match score—thus creating a "VerFlg = M"-status unlinked record.

probable links to multiple people in the reference files, and all interactions. The NAR-PIK model included the home ownership indicator, the household size indicator, an indicator for unknown relationship to the householder, the female indicator, and an indicator for the Hispanic CVAP race/ethnicity category.

The NAR-NACS group does not have ACS responses for either citizenship or the individual-level explanatory variables by construction. Therefore, we trained the model using a host of block group-level regressors as well as individual age information sourced from the NUMIDENT.[44]

## *2.6 Person-Place Modeling*

We generated person-place probabilities using the 2016-2020 ACS housing units as the test data. We used a predictive person-place model that allowed the probability that a MAFID in the AR data linked to the MAFID in the ACS data for the correct ACS year to vary across person-location based on the year they appeared in the ACS, the AR in which they appeared, the year they appeared in those AR, and the amount of time between the AR reference date and the ACS tabulation month.

The modeling was performed on a subset of records from the PPM input dataset, which was described in Section 2.2. We retained records from the PPM input dataset if the unique identifier assigned to the ACS housing unit on the record from the Census Bureau's Master Address File (MAFID) and the ACS year on the record linked to the MAFID and year of a housing unit that responded to the ACS between 2016 and 2020. We focused on PPM-input records with a corresponding ACS response, because we needed the ACS data to determine whether the person identified by the PIK on each record lived at the housing unit (MAFID) on the record in the housing unit's ACS tabulation month. This approach assumes that the characteristics that predict a person's presence in the housing unit are the same for responding and nonresponding ACS housing units.

The outcome variable for the models was an indicator variable that equaled one if the administrative PIK-MAFID pair and the ACS year on a PPM-input record matched the PIK-MAFID pair and year of a person in a housing unit that responded to the ACS between 2016 and 2020. The model predicted the likelihood the PIK on the PPM-input record was at the ACS MAFID on the record when the household submitted its ACS survey response.

We split the PPM input dataset into three subsamples: a 25 percent training sample of the ACS matched to administrative addresses for fitting the model, a 50 percent holdout sample for testing the generalizability of the model, and a 25 percent sample for model parameter tuning. We then assigned

[44] Specifically, we use the population, proportion of individuals with positive wages/salary income, proportion of individuals with positive total income, proportion of citizens, proportion of employed individuals, proportion of individuals who speak English "very well" or "well," proportion of fertile individuals, proportion of Hispanic individuals, proportion of non-Hispanic Black individuals, proportion of individuals who have changed dwellings, proportion of individuals who are married with a spouse present in the home, proportion of individuals receiving public assistance income, proportion of individuals born outside of the United States, proportion of individuals in poverty, proportion of individuals living in a single-family house, proportion of individuals who own their home (including with a mortgage), mean wages among positive earners, and mean total income among people with positive total income. All block group-level regressors for training the data come from the 2015-2019 ACS data.

predicted probabilities to the full sample. The final model was a logistic regression model with the following variables:

- Indicators for AR sources (see Table 2.1)
- A duration in days from the ACS tabulation month for both the first and last date the PIK-MAFID appeared in an AR source (compiled from individual source dates)
- Indicators for the first and last year the PIK-MAFID appeared in any source, with years spanning from 2015-2021

Table 2.1 Sources of administrative data on people linked to ACS housing units

| Source abbreviation | Description of data | Years covered |
|---|---|---|
| BOP | Data from Bureau of Prisons (BOP). Includes pre-prison addresses for people entering custody, post-prison addresses for people leaving custody, and approximate dates in custody. | 2015-2020 |
| FHA | Data from Federal Housing Administration (FHA). Includes addresses for borrowers on FHA mortgage insurance contracts. | 2015-2020 |
| HUD PICTRACS | Data from Department of Housing and Urban Development (HUD) Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS). Includes addresses for recipients of rental assistance and residents of public housing. | 2015-2021 |
| IHS | Data from Indian Health Service (IHS) Patient Registration File. Includes addresses for IHS patients. | 2015-2020 |
| IRS 1040 | Data from Internal Revenue Service (IRS) Forms 1040. Includes addresses for people listed on an individual income tax return as the primary filer, the secondary filer, or a dependent. | 2015-2021 |
| IRS 1099 | Data from Internal Revenue Service (IRS) Forms 1099. Includes addresses for people listed on an information return as the recipient of a non-employee payment such as independent contractor income, rent, royalties, interest, dividends, or gambling winnings | 2015-2021 |
| IRS 1099-R | Data from Internal Revenue Service (IRS) Forms 1099-R. Includes addresses for people listed on an information return as the recipient of a distribution from a retirement account. | 2015-2021 |
| MBR | Data from Social Security Administration (SSA) Master Beneficiary Record. Includes addresses for claimants applying for retirement, survivor, or disability benefits and for those enrolling in the Hospital Insurance or Supplementary Medical Insurance program. | 2019-2020 |
| Medicaid | Data from Centers for Medicare and Medicaid Services (CMS) Transformed Medicaid Statistical Information System (T-MSIS). Includes addresses for Medicaid beneficiaries. | 2015-2019 |
| Medicare | Data from Centers for Medicare and Medicaid Services (CMS) Medicare Enrollment Database (MEDB). Includes addresses for Medicare beneficiaries. | 2015-2021 |
| NCOA | Data from U.S. Postal Service (USPS) National Change of Address (NCOA) System. Includes “from” and “to” addresses | 2015-2021 |

| Source abbreviation | Description of data | Years covered |
|---|---|---|
| | of people reporting a temporary or permanent change of address. | |
| Passports | Data from U.S. Department of State passport records. Includes addresses of U.S. passport holders. | 2015-2020 |
| SSR | Data from Social Security Administration (SSA) Supplemental Security Record (SSR). Includes addresses for people currently or formerly eligible for Supplemental Security Income (SSI) and for their co-resident spouses and parents. Supplemental Security Income is cash assistance for people who are elderly, blind, or have a disability. | 2016, 2019-2021 |
| SSS | Data from Selective Service System (SSS) Registration File. Includes addresses of registrants. | 2015-2021 |
| USCIS | Data from U.S. Citizenship and Immigration Services (USCIS). Includes addresses for naturalized citizens and noncitizens from approved applications for naturalization, lawful permanent resident, asylee, and refugee status. | 2015-2020 |
| USMS | Data from U.S. Marshals Service (USMS) Prisoner Processing and Population Management/Prisoner Tracking System (PPM-PTS). Includes facility addresses and alternative/home addresses for people in custody and people received into custody. | 2015-2020 |
| VSGI | Data from Veterans Service Group of Illinois (VSGI) Consumer Referential Database (CRDB). Includes addresses from magazine and periodical change of address data, utility records, and other sources. | 2015-2021 |
| AK PFD | Data from Alaska Department of Revenue, Permanent Fund Division (AK PFD). Includes addresses for applicants and recipients of annual dividends paid to Alaska residents. | 2015-2020 |
| DMV | Data from driver's licenses and state identifications recently issued or updated in five states. | Varies by state |
| SNAP | Data from Supplemental Nutrition Assistance Program (SNAP) in 24 states. | Varies by state |
| TANF | Data from Temporary Assistance for Needy Families (TANF) program in 22 states. | Varies by state |
| WIC | Data from Special Supplemental Nutrition Assistance Program for Women, Infants, and Children (WIC) in 18 states. | Varies by state |

Additional variables were available but tended to be either underpopulated, such as source-specific date variables, or were correlated with the variables included, such as additional variables from the IRS 1040 tax records. All these additional variables were included in an elastic net logistic regression, which empirically chooses between the variables provided. However, adding these additional variables resulted in non-convergence in the standard logistic regression. The processing time on the elastic net prevented us from using the predicted probabilities from it, but the results will serve as a valuable comparison.

Table 2.2 Person-Place Model Validation Statistics

| Year | Within-Sample Correctly Classified using Pr(D)>=0.5 | Validation Sample median prob if acs_resp_pm==1 | Validation Sample median prob if acs_resp_pm==0 |
|---|---|---|---|
| 2016 | 81.99% | 0.7963 | 0.1174 |
| 2017 | 81.90% | 0.7938 | 0.1278 |
| 2018 | 82.41% | 0.8066 | 0.1092 |
| 2019 | 82.76% | 0.8135 | 0.1275 |
| 2020 | 83.09% | 0.8327 | 0.1180 |

Notes: The table shows the validation statistics for a sample of ACS Housing Units linked to AR PIKs at the same MAFID, with the prediction on whether the PIK was in the survey sample, by year of the ACS survey. Column 1 shows the within-sample classification error on the 25 percent training sample, and columns 2 and 3 show the median predicted probability if the dependent variable is equal to 1 (linked to the ACS survey sample) or 0 (non-linked) within the 50% holdout sample. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

We ran the model by ACS year to allow predicted probabilities to change over time. Table 2.2 gives basic validation statistics by year. The first column shows classification rates within the training dataset, with a predicted probability of greater than 0.5 modeled as a match to the ACS location. The classification rates are all above 80%. This analysis does not take into account multiple addresses per person in order to assign each person to just one location but uses each observation individually. In the holdout sample, the predicted probabilities display a separation between those that match, with predicted probabilities in the 0.79-0.83 range, and those that do not match, with predicted probabilities in the 0.11-0.13 range.

## *2.7 Tabulation*

Once all the assembly of relevant administrative records (AR) and modeling was complete, we implemented several additional steps to tabulate the data, as described here.

First, we constructed a person-level file. For people in group quarters, we used the post-disclosure avoidance 2016-2020 ACS file for the 50 states and the District of Columbia (DC).[45] Age, race/ethnicity, and as-reported citizenship came from the post-disclosure avoidance file. When citizenship was edited or imputed, we substituted AR citizenship when available, and we modeled citizenship otherwise.

For people in responding ACS housing units, we included only the housing unit CMIDs (Continuous Measurement IDs, which represent an ACS survey case) and PNUMs (person IDs) that were in the post-disclosure avoidance 2016-2020 ACS file for the 50 states and DC.[46] Their age, race/ethnicity, and as-

[45] Changing how citizenship statistics are produced for Puerto Rico is out of scope for this project. Though some AR are available for Puerto Rico, they are not nearly as complete as for the 50 states and the District of Columbia. Note that people in group quarters are included in the post-disclosure avoidance file where person data in housing units are swapped, but none of the group quarters people are swapped. Group quarters people are not included in the pre-disclosure avoidance files, and group quarters responses are never swapped.

[46] Note that an address can contain multiple CMIDs, as it may be included in the sample multiple times. There are some cases that are in an ACS panel from late 2015 with a tabulation month in 2016 that are re-sampled in 2020.

reported citizenship were taken from the pre-disclosure avoidance ACS files.[47] As with group quarters people, AR citizenship was used in place of edited or imputed ACS citizenship, when available. Otherwise, modeled citizenship (described in Section 2.5) was used when ACS citizenship was edited or imputed.

For nonresponding ACS housing units, we populated their rosters with people who have PIKs and were placed at the ACS housing unit by an administrative record with a vintage within 12 months on either side of the ACS tabulation month.[48] Children not otherwise found in AR during this window, but who were linked via the Household Composition Key to a parent who had an AR record with an ACS housing unit address in the time window, were placed in the same housing unit(s) as their parent. These children were given the same person-place probabilities as their parents.

Probabilities of being voting age, of being in each CVAP race/ethnicity group, and of being a citizen were calculated as described in Sections 2.2 and 2.3 and were attached to each PIK.

The person-place probability that the AR person resides at a particular nonresponding ACS housing unit in the tabulation month comes from the person-place model described in Section 2.6. We imposed a normalization on the probability by dividing it by the sum of all the person's person-place probabilities across their AR addresses within the 25-month window around the tabulation month. Hence, the person received a total weight of one across all in-scope AR addresses, whether they were in the initial ACS sample or not.

The normalization procedure implicitly assumes that the person's residence in the ACS tabulation month was included in their set of AR addresses. Undoubtedly, that assumption was sometimes violated in the data. Those missing their tabulation month residence were likely to have fewer AR addresses than others, in which case the normalization increased the weight on their ACS address more. At the same time, there were surely people who resided at an ACS address in the tabulation month, but that address was not among their AR addresses, resulting in an AR omission. The overweighting of people missing an AR with their residence in the tabulation month could at least partially balance these omissions.

A group of particular concern when estimating the citizen share is people who lived outside the United States in the tabulation month yet had an ACS address among their AR. People sometimes living abroad might have even fewer AR addresses than others who violated the normalization assumption above, so the normalization would increase the weight on their ACS address even more. It is probable that a high share of the people splitting time between the U.S. and other countries are noncitizens. The normalization could exacerbate the downward bias in the estimated AR citizen share caused by part-time U.S. residents.

Given this concern, we study the sensitivity of the estimates to how the person-place probability weights are constructed by producing results using alternative person-place probability weights. One

---

We treat each survey case separately – an address could have a separate household roster for each time it appears in the survey sample.

[47] This way swapping occurs only once, after the person file has been assembled.

[48] When processing administrative data that included sufficient information to assign reference years but not reference months to observations of person-housing unit pairs, we used administrative records with a vintage within one year on either side of the ACS tabulation year.

version uses the person-place probabilities produced by the model without normalization. A second version adds to the normalization denominator the probability that the person's ACS tabulation month address was missing from their AR addresses. We estimated this probability via a logistic regression with a dependent variable equal to one if the ACS person's ACS address was not included in their AR and zero otherwise. The regression sample was 2016-2020 ACS people with PIKs (and thus potentially linkable to AR). The regressors were citizenship status-age-race/ethnicity categorical variables. Here citizenship status was divided into citizens, noncitizens with SSNs, and noncitizens with ITINs.[49]

For all three of these methods, we summed the person-place probabilities across all the AR addresses for AR people having the nonresponding housing unit address, then excluded the AR for those units that had a total that exceeded the maximum for ACS housing units (i.e., the population for the unit becomes zero).[50] Such cases may have occurred when the address used in AR was not the person's residence (e.g., using the tax preparer's address on an IRS 1040 or 1099 form rather than the filer's residential address).

To provide disclosure protection, swapping was then performed on a subset of the housing units in this file,[51] using the current ACS swapping criteria modified to accommodate all units in the frame. All the people who were in a selected housing unit A were moved to housing unit B, and those initially in housing unit B were moved to housing unit A. The procedure used the current ACS swapping criteria modified to accommodate all units in the frame, including nonresponding ones.

As in the traditional ACS, the person weights were adjusted so that the estimates approximate the PEP population at the ACS estimation stratum level by age and race/ethnicity.[52] We used just two age categories, 0-17 and 18 and over, since those were the groups relevant for CVAP. The race/ethnicity categories used for the weighting were non-Hispanic American Indian or Alaska Native (AIAN), non-Hispanic Asian, non-Hispanic Black, non-Hispanic Native Hawaiian or other Pacific Islander (NHPI), non-Hispanic White, and non-Hispanic multi-race. In the ACS population adjustment procedure, people who were non-Hispanic multi-race were reassigned to one of the single-race categories. We kept non-Hispanic multi-race separate because we had many more observations to work with due to adding nonresponding housing units, and we also did not try to control to nearly as many characteristics.[53] We aggregated the county-level 2016-2020 PEP population estimates for housing units plus group quarters to the ACS estimation stratum level.[54,55] The 2016-2020 ACS group quarters population by age crossed with the six race/ethnicity categories, aggregated to the ACS estimation stratum level using the ACS

[49] Large differences in the average number of AR addresses for noncitizens with SSNs vs. ITINs shown in Table 3.11 motivate this distinction.

[50] We leave alternative solutions to this issue to future research.

[51] Here and in subsequent processing steps, we are referring to the file using person-place probabilities with our preferred normalization, not including in the denominator the probability the person's ACS tabulation month address is missing from their AR. The comparison of the different ways of handling the person-place probabilities uses pre-disclosure avoidance files.

[52] Most ACS estimation strata are individual counties. Smaller counties within the same state are combined, as is done in the ACS.

[53] The ACS population adjustments control to several age group categories, while we focus on age 0-17 and age 18 and over. Household composition (e.g., whether the householder is in a two-partner relationship or not) is also controlled in the ACS population adjustments, but not here.

[54] There are 2,988 ACS estimation strata for 3,143 counties.

[55] We divided the annual PEP estimates by five to obtain the average over the 5-year period.

final person weights,[56] was subtracted from the PEP estimates to obtain the housing unit population controls.

We aggregated people in housing units to the ACS estimation stratum level using county codes from the time of the survey (not 2020 Census tabulation geography). The person weights for each age-race/ethnicity cell were the probability of being in the age category, multiplied by the probability of being in the race/ethnicity category, multiplied by the normalized person-place probability, multiplied by the ACS housing unit base weight.

We mimicked the ACS collapsing procedures.[57] Each age-race/ethnicity weighting group must contain a minimum of 10 sample people and a population control to initial sample estimate between 1/3.5 and 3.5. Any group that violated these requirements was collapsed until all groups satisfied the criteria.

Collapsing was done in the following order:

1. *If the requirements are not met when all non-Hispanic [age-]race groups are combined then all weighting [age-]race-ethnicity groups are collapsed together and the collapsing is complete.*
2. *If the requirements are not met for Hispanics [by age group], the Hispanics [by age group] are collapsed with the largest non-Hispanic non-White [age] group.*
3. *If the requirements are not met for any non-Hispanic non-White [age] group, it is collapsed with the largest (prior to collapsing) non-Hispanic non-White [age] group.*
4. *If the largest collapsed non-Hispanic non-White group still does not meet the requirements, it is collapsed with the surviving non-Hispanic non-White groups in the following order until the requirements are met: Black, American Indian and Alaskan Native, Asian, and Native Hawaiian or Pacific Islander [, and multi-race].*
5. *If all non-Hispanic non-White [age] groups have been collapsed together and the collapsed group still does not meet the requirements, it is collapsed with the non-Hispanic White [age] group.*
6. *If the requirements are not met for the non-Hispanic White [age] group, then it is collapsed with the largest non-Hispanic non-White [age] group.*[58]

Note that we did not collapse the two age groups. Once collapsing was completed, we aggregated the PEP estimates and our CVAP estimates into the new groupings. The population adjustment factor for the collapsed age-race/ethnicity category in the ACS estimation stratum was the PEP estimate divided by our CVAP estimate (using the same person weights as those used to aggregate to the ACS estimation stratum level, again applied to the swapped person file).

Our final person-age-race/ethnicity weights were the probability of being in the age category, multiplied by the probability of being in the race/ethnicity category, multiplied by the normalized

[56] Entropy balance weights did not apply to the group quarters population.
[57] See U.S. Census Bureau (2014a), pp. 20-21.
[58] U.S. Census Bureau (2014b), p. 21. Our modifications are in brackets [].

person-place probability, multiplied by the ACS housing unit base weight, multiplied by the population adjustment factor specific to that age-race/ethnicity group in that ACS estimation stratum.[59,60]

In the tabulation we calculated the contribution of each person in a housing unit to each citizen-age-race/ethnicity cell as the probability of being a citizen, multiplied by the probability of being in the age group, multiplied by the probability of being in the race/ethnicity category, multiplied by the normalized person-place probability, multiplied by the ACS housing unit base weight, multiplied by the population adjustment factor for that age-race/ethnicity group. For group quarters people, it was the probability of being a citizen, multiplied by the probability of being in the age group, multiplied by the probability of being in the race/ethnicity category, multiplied by the normalized person-place probability, multiplied by the 2016-2020 ACS final person weight. People who came from ACS responses were given a normalized person-place probability equal to one, both for those in housing units and group quarters. Estimates for each level of geography (national, state, county, tract, block group, Minor Civil Division, congressional district, state legislative district upper chamber, and state legislative district lower chamber) used 2020 tabulation geography.

Margins of error (MOEs) were calculated by first producing replicate population adjustment factors as described above, but rather than using ACS housing unit base weights, base replicate weights (of which there are 80 separate ones for each housing unit) were used. Then we performed the same tabulation procedure as above 80 times, substituting replicate base weights for the base weights and replicate population adjustment factors for the population adjustment factors used in the main estimates. The variance $v(\hat{\theta}_0)$ was calculated using the formula:

$$v(\hat{\theta}_0) = \frac{4}{80}\sum_{r=1}^{80}(\hat{\theta}_r - \hat{\theta}_0)^2$$

where $\hat{\theta}$ is an estimate, $\hat{\theta}_0$ is the estimate computed using the main estimation weights, and $\hat{\theta}_r$ are replicate estimates (r=1, …., 80).[61] The MOE is:

$$MOE(\hat{\theta}_0) = 1.645 \times \sqrt{v(\hat{\theta}_0)}$$

The MOEs can be used to create 90 percent confidence intervals around the estimates.

[59] In cases where multiple age-race/ethnicity groups had been collapsed to calculate the population adjustment factor, the same factor was applied to each of them. When a characteristic came from an ACS response, the probability that the person has that characteristic was set equal to one.

[60] Note that this weighting did not include the other adjustments that were made to the traditional and entropy balance weights used in the ACS, such as subcounty and housing unit controls, or family equalization. These adjustments may be correlated with citizenship (e.g., if weights are redistributed from other household members to the householder, and the householder has a different propensity to be a citizen than other members) and could potentially contribute to the differences in citizenship estimates.

[61] See U.S. Census Bureau, (2014c), pp. 4-5.

In cases where the estimate was fully controlled to the PEP estimates, the standard errors and thus the MOEs were set to zero.[62] For estimates equal to zero, we used the MOEs produced by the ACS CVAP tabulations using entropy balance weights.[63]

Finally, ACS rounding rules were applied to all estimates for additional disclosure protection.

# 3 Results

There are two main differences in how the ACS and enhanced CVAP tabulations are constructed. The first difference is regarding the citizenship variable itself. The ACS CVAP uses ACS citizenship as reported on the survey when it is available and otherwise uses edited or imputed citizenship. The enhanced CVAP uses as-reported ACS citizenship for people in responding ACS housing units when available. For people in responding ACS housing units whose ACS citizenship is edited or imputed, the enhanced CVAP brings in citizenship information from administrative records (AR) or modeling. For nonresponding housing units, enhanced CVAP uses AR citizenship when available and modeled predictions otherwise. The second difference is that ACS CVAP adjusts the weights in the responding units to account for nonresponse bias and the sampling for the CAPI mode of data collection, while enhanced CVAP uses AR to create response records for the nonresponding units.[64] We study these two differences separately to understand the extent to which the overall observed differences are due to each of these changes.[65]

## *3.1 Citizenship Variable Comparison*

We first study how the average citizen share varies across sources in different parts of the initial ACS sample and the importance of each group. In Table 3.1 we divide the ACS units into group quarters (GQ), responding housing units, and nonresponding housing units. We further divide each group into subgroups depending on which citizen values are available. The ACS citizenship variable types are as-reported, edited, and imputed values. The AR types are either non-missing AR citizenship (referred to simply as AR citizenship) or modeled citizenship. Here the only difference between the enhanced and AR citizen shares is that the enhanced column uses as-reported ACS values when available.[66] The percentages are weighted by final person weights for group quarters people[67] and base weights for people in housing units. The latter allows us to focus on the citizen values, abstracting from differences in weight adjustments between the methods, which will be examined in Sections 3.2-3.4.

---

[62] We approximate this by setting the MOE to zero when the estimate is within 5 of the PEP estimate.

[63] See U.S. Census Bureau (2014c), pp. 5-6.

[64] The enhanced CVAP nonresponse methodology fills each housing unit with people with PIKs from AR who have that address within one year of the ACS tabulation date. The probability that the person is at the ACS address in the ACS tabulation month is modeled. Age, race/ethnicity, and citizenship are sourced from past Title 13 surveys and administrative records, when available, and we use modeled probabilities otherwise.

[65] Note also that it is possible to use one approach for citizenship item nonresponse, and the other approach to address unit nonresponse.

[66] We use the ACS household roster for all three columns here, whereas an AR roster is used in the AR column in some analyses below.

[67] There are no analogous base weights for group quarters people. The pre-disclosure avoidance file does not contain imputed person records, so the number of person records is different.

GQs make up a small share the ACS person sample,[68] and the citizen shares are very similar across sources. Since the bulk of the differences come from housing units, we exclude GQs from the exercises in Sections 3.2 and 3.3.

Table 3.1 Percent Citizens in 2016-2020 ACS, Enhanced CVAP, and AR

| | ACS % Citizens | Enhanced % Citizens | AR % Citizens | % of Sample |
|---|---|---|---|---|
| GQ, As-Reported ACS, AR | 97.34 | 97.34 | 97.28 | 0.66 |
| GQ, As-Reported ACS, Modeled | 94.05 | 94.05 | 93.85 | 1.77 |
| GQ, Edited ACS, AR | 97.73 | 98.19 | 98.19 | 0.00 |
| GQ, Edited ACS, Modeled | 96.31 | 90.94 | 90.94 | 0.01 |
| GQ, Imputed ACS, AR | 93.58 | 95.88 | 95.88 | 0.01 |
| GQ, Imputed ACS, Modeled | 90.83 | 91.66 | 91.66 | 0.16 |
| GQ Total | 94.69 | 94.74 | 94.59 | 2.62 |
| | | | | |
| HU, As-Reported ACS, AR | 95.95 | 95.95 | 95.89 | 50.84 |
| HU, As-Reported ACS, Modeled | 81.47 | 81.47 | 81.63 | 4.18 |
| HU, Edited ACS, AR | 91.13 | 91.48 | 91.48 | 0.17 |
| HU, Edited ACS, Modeled | 75.36 | 79.10 | 79.10 | 0.03 |
| HU, Imputed ACS, AR | 94.12 | 92.93 | 92.93 | 3.33 |
| HU, Imputed ACS, Modeled | 88.59 | 81.64 | 81.64 | 0.83 |
| Nonresponding HU, AR | N.A. | 89.59 | 89.59 | 37.91 |
| Nonresponding HU, Modeled | N.A. | 89.82 | 89.82 | 0.09 |
| Responding HU Total | 94.70 | 94.54 | 94.50 | 59.38 |
| Nonresponding HU Total | N.A. | 89.60 | 89.60 | 38.00 |
| HU Total | 94.70 | 92.61 | 92.58 | 97.38 |
| Total | 94.70 | 92.66 | 92.64 | 100.0 |

Notes: The household rosters are from the pre-disclosure avoidance 2016-2020 ACS file for housing units and the post-disclosure avoidance 2016-2020 ACS file for group quarters in all columns.[69] The group quarters person observations are weighted by final person weights, and the housing unit person observations are weighted by housing unit base weights. ACS citizenship comes from the pre-disclosure avoidance 2016-2020 ACS file for housing units and the post-disclosure avoidance 2016-2020 ACS file for group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

About half of the sample is people in responding housing units with as-reported ACS citizenship and AR citizenship. About 6.1 percent of people classified as noncitizens in AR are reported as citizens in the

[68] The sample shares of group quarters relative to housing units are not completely comparable in Table 3.1, since final weights are used for people in group quarters and base weights are used for those in housing units. The published estimates using final person weights for people in housing units and group quarters show a group quarters share of 2.5 percent rather than 2.6 percent.

[69] The pre-disclosure file for housing units is also called the unswapped file, and the post-disclosure file for housing units and group quarters is also called the swapped file. Note, however, that swapping is only applied to housing units. A different disclosure avoidance technique called partial synthesis is applied to group quarters.

ACS, while just 0.21 percent of those classified as citizens in AR are reported as noncitizens in the ACS (Table 3.2B). The patterns are similar when starting with the ACS response: 4.75 percent of ACS citizens are AR noncitizens, while just 0.27 percent of ACS noncitizens are AR citizens (Table 3.2B).[70] The higher share of AR noncitizens classified as ACS citizens compared to ACS noncitizens classified as AR citizens suggests that the ACS citizenship responses by AR noncitizens may reflect outdated AR citizenship information or social desirability bias in ACS citizenship responses. The citizen share from both sources is virtually identical (95.95 percent in the ACS and 95.89 percent in AR in Table 3.1), however, so the choice between them has little effect overall.

When using edited ACS citizenship, the citizen share is a bit lower than AR or modeled citizenship, whereas the share is higher when using imputed ACS citizenship, which justifies distinguishing edits from imputations in our analysis. Edited cases make up a tiny fraction of the sample, however, so it is difficult to draw strong conclusions about them. The citizen share is nearly 7 percentage points higher with imputed than modeled citizenship. Both estimates in the "HU, Imputed ACS, AR" row are higher than each estimate in the "HU, Imputed ACS, Modeled" row. We will explore these patterns further below.

The citizen share is nearly 5 percentage points lower in nonresponding housing units than responding ones, and nonresponding units make up 38 percent of the sample (Table 3.1). This difference accounts for most of the overall citizen share difference between the ACS and the estimates from enhanced CVAP and AR. Below we examine how much this gap changes when applying the final ACS person weights.

Table 3.2A Comparison of All ACS vs. AR/Modeled Citizenship Values

| | AR/Modeled Noncitizen | AR/Modeled Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS Noncitizen | 4.30 | 1.20 | 5.50 |
| ACS Citizen | 1.40 | 93.10 | 94.50 |
| AR Total | 5.70 | 94.30 | 100.0 |
| | Column Percentages | | |
| ACS Noncitizen | 75.48 | 1.28 | |
| ACS Citizen | 24.52 | 98.72 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS Noncitizen | 78.12 | 21.88 | 100.0 |
| ACS Citizen | 1.48 | 98.52 | 100.0 |

Notes: N=23,330,000 observations for all people in the 2016-2020 ACS. AR citizenship is merged in by PIK. The observations are weighted by enhanced CVAP population-adjusted person weights (194,200,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

[70] Jasso and Rosenzweig (2020)) suggest that a reason for citizens being misreported as noncitizens is that people receiving citizenship through a parent's naturalization (derived citizenship) do not fit neatly into the ACS citizen categories, since they were neither born in the U.S. nor naturalized. The results in Table 3.2B suggest that this happens infrequently. Note also that record linkage error could account for some of the discrepancy.

Table 3.1 shows average differences across sources, while Tables 3.2A-J also display the share of records whose citizenship values disagree. The latter could be important for statistics at lower levels of geography, as discrepancies may not average out as much there. Table 3.2A compares ACS citizenship that is either as-reported, edited, or imputed to AR or modeled citizenship among people in ACS responding unit rosters.[71] The cell percentages are the shares of all people in the table in each ACS-AR citizenship combination. The column percentages reflect shares of people with a particular AR citizenship status who are noncitizens or citizens according to the ACS citizenship variable. The row percentages show the shares of people with a particular ACS citizenship status who are AR noncitizens vs. citizens. The cell percentages part of the table shows that the total citizen share is 0.2 percentage points higher when using ACS citizenship than AR citizenship (94.5 vs. 94.3 percent). The two sources disagree in 2.6 percent of cases (adding the two off-diagonal cells in the cell percentages part of the table).

Table 3.2B Comparison of As-Reported ACS vs. AR Citizenship

| | AR Noncitizen | AR Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 4.02 | 0.20 | 4.22 |
| ACS As-Reported Citizen | 0.26 | 95.52 | 95.78 |
| AR Total | 4.28 | 95.72 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 93.89 | 0.21 | |
| ACS As-Reported Citizen | 6.11 | 99.79 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS As-Reported Noncitizen | 95.26 | 4.75 | 100.0 |
| ACS As-Reported Citizen | 0.27 | 99.73 | 100.0 |

N=19,080,000 observations for all people with as-reported ACS citizenship and AR citizenship merged in by PIK. The observations are weighted by enhanced CVAP population-adjusted person weights (154,800,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

We focus on as-reported ACS vs. AR citizenship in Table 3.2B. Here the total citizen shares are nearly identical (0.06 percent different), and the disagreement rate is just 0.46 percent. It implies that replacing as-reported ACS with AR citizenship is likely to have a minimal effect on the estimates even at lower levels of geography. It also suggests that record linkage errors between AR and the ACS do not exhibit citizenship bias. Hereafter, we use as-reported ACS citizenship as a comparator for the modeling used in enhanced CVAP and AR citizenship as the comparator for ACS edits and imputations.[72]

[71] Note that we use AR citizenship here in place of as-reported ACS citizenship just for evaluation purposes. As-reported ACS citizenship is used when available in both the ACS and enhanced CVAP tabulations.

[72] This use is justifiable, if AR and as-reported ACS citizenship are accurate even when the other source is not available.

Table 3.2C Comparison of Edited or Imputed ACS vs. AR Citizenship

| | AR Noncitizen | AR Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS Edited/Imputed Noncitizen | 2.46 | 3.76 | 6.21 |
| ACS Edited/Imputed Citizen | 4.86 | 88.93 | 93.79 |
| AR Total | 7.31 | 92.69 | 100.0 |
| | Column Percentages | | |
| ACS Edited/Imputed Noncitizen | 33.59 | 4.06 | |
| ACS Edited/Imputed Citizen | 66.41 | 95.95 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS Edited/Imputed Noncitizen | 39.52 | 60.48 | 100.0 |
| ACS Edited/Imputed Citizen | 5.18 | 94.82 | 100.0 |

N=1,275,000 observations for people with edited or imputed ACS citizenship and AR citizenship. AR citizenship is merged in by PIK. The observations are weighted by enhanced CVAP population-adjusted person weights (11,050,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.2D Comparison of Edited ACS vs. AR Citizenship

| | AR Noncitizen | AR Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS Edited Noncitizen | 3.15 | 5.97 | 9.12 |
| ACS Edited Citizen | 5.38 | 85.50 | 90.88 |
| AR Total | 8.52 | 91.48 | 100.0 |
| | Column Percentages | | |
| ACS Edited Noncitizen | 36.92 | 6.53 | |
| ACS Edited Citizen | 63.08 | 93.47 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS Edited Noncitizen | 34.50 | 65.50 | 100.0 |
| ACS Edited Citizen | 5.92 | 94.08 | 100.0 |

N=64,000 observations for people with edited ACS citizenship and AR citizenship. AR citizenship is merged in by PIK. The observations are weighted by enhanced CVAP population-adjusted person weights (549,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.2E Comparison of Imputed ACS vs. AR Citizenship

| | AR Noncitizen | AR Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS Imputed Noncitizen | 2.42 | 3.64 | 6.06 |
| ACS Imputed Citizen | 4.83 | 89.11 | 93.94 |
| AR Total | 7.25 | 92.75 | 100.0 |

| | Column Percentages | | |
|---|---|---|---|
| ACS Imputed Noncitizen | 33.38 | 3.93 | |
| ACS Imputed Citizen | 66.62 | 96.07 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS Imputed Noncitizen | 39.92 | 60.08 | 100.0 |
| ACS Imputed Citizen | 5.14 | 94.86 | 100.0 |

N=1,211,000 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (10,500,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Tables 3.2C-E show ACS citizenship edits and imputations compared to AR citizenship. Among the ACS values that are not as-reported, edited values have a higher disagreement rate (11.35 vs. 8.47 percent), while imputations are more different on average (1.19 vs. 0.6 percentages points different from AR).

Table 3.2F Comparison of As-Reported ACS vs. Modeled Citizenship

| | Modeled Noncitizen | Modeled Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 3.55 | 5.50 | 9.05 |
| ACS As-Reported Citizen | 5.52 | 85.44 | 90.95 |
| AR Total | 9.06 | 90.94 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 39.16 | 6.05 | |
| ACS As-Reported Citizen | 60.84 | 93.95 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS As-Reported Noncitizen | 39.22 | 60.78 | 100.0 |
| ACS As-Reported Citizen | 6.06 | 93.94 | 100.0 |

N=951,000 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (6,294,000 weighted people). Observations included in the training for the NAR-PIK, NAR-SS, or NAR-NSS models are excluded from the analysis. The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Modeled estimates are virtually identical on average compared to as-reported ACS values (see Table 3.2F), as should be the case when subsets of ACS observations are used to train the models. This contrasts with ACS edits and imputations, which show a difference of 1.1 percentage points from AR citizenship in Table 3.2C. About 11 percent of modeled observations disagree with as-reported ACS values, a rate falling in between the imputation and edit disagreement rates in Tables 3.2D-E.

Table 3.2G Comparison of ACS to Enhanced Citizenship, As-Reported ACS vs. Modeled NAR-PIK

| | Modeled NAR-PIK Noncitizen | Modeled NAR-PIK Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 1.42 | 7.05 | 8.46 |
| ACS As-Reported Citizen | 6.25 | 85.29 | 91.54 |
| AR Total | 7.67 | 92.33 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 18.46 | 7.63 | |
| ACS As-Reported Citizen | 81.54 | 92.37 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS As-Reported Noncitizen | 16.73 | 83.27 | 100.0 |
| ACS As-Reported Citizen | 6.83 | 93.17 | 100.0 |

N=199,000 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (1,562,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.2H Comparison of ACS to Enhanced Citizenship, As-Reported ACS vs. Modeled NAR-SS

| | Modeled NAR-SS Noncitizen | Modeled NAR-SS Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 15.93 | 10.35 | 26.28 |
| ACS As-Reported Citizen | 10.26 | 63.46 | 73.72 |
| AR Total | 26.19 | 73.81 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 60.83 | 14.03 | |
| ACS As-Reported Citizen | 39.17 | 85.97 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS As-Reported Noncitizen | 60.62 | 39.38 | 100.0 |
| ACS As-Reported Citizen | 13.92 | 86.08 | 100.0 |

N=121,000 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (1,008,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.2I Comparison of ACS to Enhanced Citizenship, As-Reported ACS vs. Modeled NAR-NSS

| | Modeled NAR-NSS Noncitizen | Modeled NAR-NSS Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 1.31 | 2.81 | 4.12 |
| ACS As-Reported Citizen | 3.39 | 92.50 | 95.88 |
| AR Total | 4.70 | 95.30 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 27.88 | 2.95 | |
| ACS As-Reported Citizen | 72.12 | 97.05 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |

| | | | |
|---|---|---|---|
| ACS As-Reported Noncitizen | 31.81 | 68.19 | 100.0 |
| ACS As-Reported Citizen | 3.53 | 96.47 | 100.0 |

N=1,100 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (9,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.2J Comparison of As-Reported ACS vs. Modeled NAR-NPVS

| | Modeled NAR-NPVS Noncitizen | Modeled NAR-NPVS Citizen | ACS Total |
|---|---|---|---|
| | Cell Percentages | | |
| ACS As-Reported Noncitizen | 1.09 | 3.54 | 4.63 |
| ACS As-Reported Citizen | 3.92 | 91.44 | 95.37 |
| AR Total | 5.02 | 94.98 | 100.0 |
| | Column Percentages | | |
| ACS As-Reported Noncitizen | 21.77 | 3.73 | |
| ACS As-Reported Citizen | 78.23 | 96.27 | |
| AR Total | 100.0 | 100.0 | |
| | Row Percentages | | |
| ACS As-Reported Noncitizen | 23.58 | 76.42 | 100.0 |
| ACS As-Reported Citizen | 4.11 | 95.89 | 100.0 |

N=629,000 person observations. The observations are weighted by enhanced CVAP population-adjusted person weights (3,715,000 weighted people). The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Tables 3.2G-J display citizenship modeling relative to the as-reported ACS benchmark among observations for which the models are designed and which have as-reported ACS citizenship. The modeled disagreement rate is high for the group sent to PVS search but without a PIK (NAR-SS), at 20.61 percent, as well as for those with PIKs and no AR citizenship (NAR-PIK), at 13.3 percent. The disagreement rates for the other modeled groups are lower than they are for ACS edits or imputations, at 7.46 percent for those in group quarters but not part of the PVS process at all (NAR-NPVS), and 6.2 percent for those not sent to PVS search (NAR-NSS). None of the models have a mean difference with the benchmark as high as ACS imputations, at 0.79 percentage points for NAR-PIK, 0.09 percentage points for NAR-SS, 0.58 percentage points for NAR-NSS, and 0.39 percentage points for NAR-NPVS.

The results show that the average difference is generally higher in ACS edits and imputations than in the models. Table 3.3 explores this in more depth, focusing on the NAR-SS and NAR-NSS groups. To evaluate the value of using separate models for the two groups, we compare the models' predictions to as-reported ACS citizenship, among people not used in model training who have as-reported ACS citizenship and are in the NAR-SS group. We also do this comparison substituting the NAR-NSS group for the NAR-SS group. In addition, we compare the two models' predictions to edited or imputed ACS citizenship among people not included in model training and with edited or imputed ACS citizenship and who are in the NAR-SS group. We repeat this for the NAR-NSS group.

Table 3.3 Percent Citizens by ACS Reporting Type and Model Type

| | ACS As Reported Citizenship | |
|---|---|---|
| | NAR-SS | NAR-NSS |
| ACS Value | 73.72 | 95.88 |
| NAR-SS Model | 73.81 | 91.25 |
| NAR-NSS Model | 88.08 | 95.30 |
| N | 121,000 | 1,100 |
| | ACS Edited or Imputed Citizenship | |
| ACS Value | 85.86 | 94.60 |
| NAR-SS Model | 75.83 | 87.95 |
| NAR-NSS Model | 88.33 | 93.87 |
| N | 213,000 | 19,000 |

Notes: NAR-SS means the person does not have AR citizenship and was sent to PVS search but did not receive a PIK. NAR-NSS means the person does not have AR citizenship and does not have sufficient PII to be sent to PVS search. Observations included in the training for the NAR-SS or NAR-NSS models are excluded from the analysis. The observations are weighted by enhanced CVAP population-adjusted person weights. The pre-disclosure avoidance 2016-2020 ACS file is used for people in housing units, and the post-disclosure avoidance 2016-2020 ACS file is used for people in group quarters here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The as-reported ACS citizen share is much higher in the NAR-NSS group than the NAR-SS group (95.88 percent vs. 73.72 percent - see Table 3.3). The NAR-SS model produces a very similar citizen share to as-reported ACS for the NAR-SS group, and the NAR-NSS model citizen share is close to the as-reported ACS share for the NAR-NSS group. The difference between the citizen shares for the NAR-NSS model applied to the NAR-NSS group and the NAR-SS model applied to the NAR-SS group is 21.49 percentage points, similar to the 22.16 percentage point difference between as-reported ACS citizen shares. Neither model produces citizen shares close to the ACS share for the group it is not designed to model. Among observations where the ACS citizenship is edited or imputed, the NAR-SS and NAR-NSS models exhibit similar patterns to the ones for the as-reported ACS citizenship observations (18.04 percentage points higher for the NAR-NSS model when applied to the NAR-NSS group compared to the NAR-SS model applied to the NAR-SS group). The ACS edited or imputed citizen shares exhibit a much smaller difference between the NAR-SS and NAR-NSS observations, however (8.74 percentage points higher for the NAR-NSS group). *The models thus follow the as-reported ACS citizenship patterns more closely than ACS edits and imputations do.*

In sum, the NAR-SS model estimates produce smaller mean differences, but higher discrepancy rates compared to imputed ACS citizenship. The NAR-NSS and NAR-NPVS models also produce a smaller mean difference than ACS edits or imputations, and the NAR-PIK model generates a smaller mean difference than ACS imputations. ACS edits show slightly lower mean differences and discrepancies relative to the NAR-PIK model. The NAR-SS model and ACS edits have similar mean differences and discrepancies.

## *3.2 Comparing Methods in Responding and Nonresponding Housing Units in 2020*

This section examines how the enhanced CVAP methodology for nonresponding ACS housing units (called AR here, since it uses administrative records throughout) performs when estimating the population and its age and race/ethnicity characteristics. We also compare how the ACS with entropy balance weights and AR characterize the population of nonresponding ACS housing units. We focus on 2020, so we can compare the statistics to those generated from 2020 Census values for the same housing units (hereafter called 2020 Census-BW, since ACS base weights are applied when calculating the aggregate statistics). It is also an important year for nonresponding units, since they are a far higher share of units in 2020 than they were in the other years. The ACS with traditional weights is added as a comparator here, to see what effect adjustment of ACS person weights alone has on the 2020 estimates.

Though person weights can be adjusted so that CVAP population, age, and race/ethnicity estimates approximate those from population controls,[73] it is still of interest to study how well a set of estimates reflects the population prior to making those adjustments to evaluate the quality of the inputs. The population controls do not go down to the lowest geographic levels (for enhanced CVAP the population controls are at the county level, for example), so a method that performs well without the aid of adjustments is likely to perform well at lower levels of geography. Second, there is no population adjustment for citizenship. Though race/ethnicity and citizenship are highly correlated, it is not perfect, so race/ethnicity population controls are unlikely to bring citizenship estimates fully in alignment with the true values for the population. Accurate representation of population, age, and race/ethnicity characteristics prior to adjustment should increase the likelihood that a method will also do a good job with citizenship.[74]

Table 3.4A Population by Age and Race/Ethnicity, 2020 ACS Responding Housing Units

| | 2020 Census, BW | ACS, BW | AR, BW |
|---|---|---|---|
| Total Population | 129,900,000 | 130,400,000 | 133,900,000 |
| | Percent by Age | | |
| Age 0-17 | 20.32 | 20.38 | 21.84 |
| Age 18+ | 79.68 | 79.60 | 78.15 |
| | Percent of Age 0-17 | | |
| Non-Hispanic | 79.76 | 79.50 | 79.07 |
| NH AIAN Alone | 0.82 | 0.80 | 1.13 |
| NH Asian Alone | 6.41 | 6.28 | 4.66 |
| NH Black Alone | 9.38 | 8.81 | 9.71 |
| NH NHPI Alone | 0.21 | 0.18 | 0.18 |
| NH White Alone | 56.74 | 56.73 | 60.00 |
| NH AIAN & White | 1.30 | 1.47 | 0.16 |
| NH Asian & White | 2.10 | 2.23 | 1.22 |

[73] None of the estimates exactly match the population controls, due to collapsing some of the age, race, and ethnicity categories.

[74] Design-based estimation of population sizes from the ACS is best evaluated at the sampled HU level, making these comparisons particularly salient independent of the citizenship data.

| | | | |
|---|---|---|---|
| NH Black & White | 1.91 | 2.04 | 1.39 |
| NH AIAN and Black | 0.12 | 0.10 | 0.03 |
| NH Other Multi-Race | 0.77 | 0.85 | 0.60 |
| Hispanic | 20.24 | 20.50 | 20.93 |
| | Percent of Age 18+ | | |
| Non-Hispanic | 87.35 | 87.09 | 86.59 |
| NH AIAN Alone | 0.61 | 0.60 | 0.85 |
| NH Asian Alone | 6.50 | 6.50 | 5.74 |
| NH Black Alone | 8.49 | 8.11 | 8.72 |
| NH NHPI Alone | 0.15 | 0.14 | 0.15 |
| NH White Alone | 69.10 | 69.08 | 69.81 |
| NH AIAN & White | 1.09 | 1.20 | 0.14 |
| NH Asian & White | 0.57 | 0.59 | 0.47 |
| NH Black & White | 0.42 | 0.45 | 0.32 |
| NH AIAN and Black | 0.05 | 0.09 | 0.03 |
| NH Other Multi-Race | 0.35 | 0.34 | 0.37 |
| Hispanic | 12.68 | 12.89 | 13.41 |

Notes: The samples are people in housing units that are responding units in the 2020 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The 2020 Census here means 2020 Census Edited File response data for the responding 2020 ACS housing units. The pre-disclosure avoidance ACS file is used in column 2. The weights are multiplied by five when calculating total population, since the weights are for 5-year rather than 1-year estimates. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.4A shows the total population and the population shares by age and race/ethnicity for non-voting-age and voting-age separately, among responding ACS housing units.[75] Initial ACS housing unit sample base weights are used for all estimates in the table. The ACS with traditional weights and ACS with entropy balance weights methods are the same except for their weights;[76] their estimates are identical when using base weights, which is why they are represented by the same column in Table 3.4A. In the AR approach, people from administrative records are used to form the household rosters, and age and race/ethnicity come from past survey and administrative data and modeling. Enhanced CVAP uses the AR approach only for nonresponding units, but we apply the AR approach here to units that responded to the ACS, to see how it compares to the other methods for the same set of housing units.

The ACS estimates are similar in total population and age and race/ethnicity shares to the 2020 Census throughout the table, with the exception that they yield lower NH Black shares for children. The closeness of the estimates is to be expected, given that both use similar data collection procedures. AR produces a slightly higher total population estimate, as well as higher shares for children, non-Hispanic AIAN, non-Hispanic Black, non-Hispanic White, and Hispanic. The non-Hispanic Asian and non-Hispanic

[75] The original weights are multiplied by five when calculating total population in Tables 3.4A, 3.4C, and 3.4E, since the original weights are for a 5-year rather than 1-year period and would thus produce about one-fifth of the total population without adjustment.

[76] Note that the entropy balance weights are incorporated into the processing of the 2020 data only. Traditional weights are used when processing the 2016-2019 data.

multi-race shares are lower. Similarly, Brown et al. (2023) report higher AR estimates for non-Hispanic AIAN, non-Hispanic Black, and Hispanic and lower shares for non-Hispanic Asian and non-Hispanic multi-race compared to the 2020 Census as a whole. They find that most of these differences can be explained by different reporting or imputation for the same people rather than differing coverage, though a significant Hispanic coverage difference remains.

The results by age and race/ethnicity for the ACS and AR in 2016-2019 are shown in Table 3.4B. The ACS voting-age non-Hispanic Asian share is much higher in 2020 than in previous years, while the AR share is similar across years. The voting-age multi-race share is only about 0.1 percentage points higher in the ACS than AR in 2016-2019, while it is 1.35 percentage points higher in 2020, the year when the questionnaire changed to give greater opportunity to report multiple races (Marks and Rios-Vargas (2021)). The AR voting-age Hispanic share is 1.6 to 2.0 percentage points higher than the ACS one in 2016-2019, compared to 0.52 higher in 2020.

ACS CAPI responses are for households that did not self-respond and were randomly selected for CAPI follow-up interviewing. They should be most like the nonresponding households, most of which were randomly selected to not have CAPI follow-up interviewing. A comparison between ACS and AR data in the CAPI households could thus be most relevant for understanding what differences result from using AR in nonresponding housing units rather than ACS responses if they existed. Tables 3.4C-D makes the ACS-AR comparison in housing units with CAPI responses. The differences between the ACS and AR are generally larger among CAPI responses, with AR showing greater representation of people who are not non-Hispanic White. The differences narrow over time, though. The ACS initially has a much higher voting-age non-Hispanic White share, but it is lower in 2020 (4.36, 4.23, 3.47, 3.06, and -0.32 percentage points difference between the ACS and AR in 2016, 2017, 2018, 2019, and 2020, respectively). In contrast, the AR-ACS difference in voting-age Hispanic shares declines each year (by 3.60, 3.36, 2.58, 1.78, and 0.33 percentage points in 2016, 2017, 2018, 2019, and 2020, respectively). The voting-age non-Hispanic White share is also several percentage points higher and the voting-age Hispanic share is much lower in 2020 compared to previous years, whether one uses the ACS or AR data. These patterns may be related to the fact that CAPI interviews were conducted by telephone to a much greater extent in 2020 than in previous years. Shin (2021) mentions that the housing units where phone numbers could be obtained may differ systematically from housing units where they could not be obtained. About 40 percent of the CAPI workload could not be contacted. Our results are consistent with non-Hispanic Whites having a greater propensity to be contacted by telephone.

Table 3.4B Population by Age and Race/Ethnicity, 2016-2019 ACS Responding Housing Units

| | 2016 ACS, BW | 2016 AR, BW | 2017 ACS, BW | 2017 AR, BW | 2018 ACS, BW | 2018 AR, BW | 2019 ACS, BW | 2019 AR, BW |
|---|---|---|---|---|---|---|---|---|
| Total Population | 201,600,000 | 211,100,000 | 195,900,000 | 204,600,000 | 196,800,000 | 204,400,000 | 191,400,000 | 197,800,000 |
| | | | | | | | | |
| Age 0-17 | 21.50 | 23.59 | 21.52 | 23.61 | 21.16 | 22.85 | 20.66 | 22.20 |
| Age 18+ | 78.50 | 76.41 | 78.48 | 76.39 | 78.83 | 77.15 | 79.34 | 77.81 |
| | Age 0-17 | | | | | | | |
| Non-Hispanic | 80.24 | 77.88 | 80.14 | 77.74 | 80.30 | 77.94 | 80.97 | 78.69 |
| NH AIAN Alone | 0.85 | 1.11 | 0.84 | 1.12 | 0.83 | 1.12 | 0.77 | 1.09 |
| NH Asian Alone | 5.47 | 4.71 | 5.81 | 4.95 | 5.70 | 4.96 | 5.88 | 5.14 |
| NH Black Alone | 9.78 | 11.23 | 9.37 | 10.77 | 9.07 | 10.44 | 8.70 | 10.05 |
| NH NHPI Alone | 0.17 | 0.18 | 0.16 | 0.18 | 0.15 | 0.17 | 0.16 | 0.17 |
| NH White Alone | 59.16 | 56.39 | 59.03 | 56.41 | 59.45 | 56.77 | 60.28 | 57.65 |
| NH AIAN & White | 0.61 | 0.34 | 0.61 | 0.35 | 0.58 | 0.32 | 0.52 | 0.31 |
| NH Asian & White | 1.66 | 1.48 | 1.72 | 1.52 | 1.87 | 1.65 | 1.98 | 1.74 |
| NH Black & White | 1.79 | 1.69 | 1.84 | 1.70 | 1.91 | 1.73 | 1.96 | 1.80 |
| NH AIAN & Black | 0.08 | 0.05 | 0.08 | 0.05 | 0.07 | 0.05 | 0.06 | 0.04 |
| NH Other Multi-Race | 0.67 | 0.71 | 0.67 | 0.72 | 0.67 | 0.72 | 0.65 | 0.70 |
| Hispanic | 19.76 | 22.12 | 19.86 | 22.26 | 19.70 | 22.06 | 19.03 | 21.31 |
| | Age 18+ | | | | | | | |
| Non-Hispanic | 88.03 | 86.02 | 87.87 | 85.86 | 87.81 | 86.01 | 88.11 | 86.55 |
| NH AIAN Alone | 0.61 | 0.87 | 0.61 | 0.88 | 0.59 | 0.88 | 0.54 | 0.86 |
| NH Asian Alone | 5.76 | 5.59 | 6.00 | 5.81 | 5.99 | 5.83 | 6.14 | 5.97 |
| NH Black Alone | 8.78 | 9.50 | 8.49 | 9.24 | 8.20 | 8.97 | 7.94 | 8.73 |
| NH NHPI Alone | 0.13 | 0.15 | 0.13 | 0.15 | 0.13 | 0.15 | 0.13 | 0.14 |
| NH White Alone | 71.25 | 68.51 | 71.09 | 68.33 | 71.34 | 68.73 | 71.81 | 69.36 |
| NH AIAN & White | 0.48 | 0.29 | 0.45 | 0.28 | 0.43 | 0.26 | 0.38 | 0.23 |
| NH Asian & White | 0.42 | 0.42 | 0.45 | 0.44 | 0.48 | 0.47 | 0.51 | 0.50 |
| NH Black & White | 0.27 | 0.29 | 0.29 | 0.30 | 0.32 | 0.32 | 0.33 | 0.34 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NH AIAN & Black | 0.07 | 0.05 | 0.06 | 0.05 | 0.06 | 0.04 | 0.05 | 0.04 |
| NH Other Multi-Race | 0.27 | 0.37 | 0.28 | 0.37 | 0.28 | 0.37 | 0.28 | 0.37 |
| Hispanic | 11.96 | 14.00 | 12.15 | 14.16 | 12.19 | 13.99 | 11.88 | 13.46 |

Notes: The samples are people in housing units that are responding units in the 2016-2019 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS files are used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.4C Population by Age and Race/Ethnicity, 2020 ACS Housing Units with CAPI Responses

| | 2020 Census, BW | ACS, BW | AR, BW |
|---|---|---|---|
| Total Population | 24,990,000 | 22,110,000 | 25,730,000 |
| | Percent by Age | | |
| Age 0-17 | 23.09 | 23.16 | 25.96 |
| Age 18+ | 76.91 | 76.84 | 74.04 |
| | Percent of Age 0-17 | | |
| Non-Hispanic | 68.82 | 66.84 | 68.11 |
| NH AIAN Alone | 2.05 | 2.11 | 2.27 |
| NH Asian Alone | 3.96 | 3.62 | 2.60 |
| NH Black Alone | 16.73 | 15.39 | 16.54 |
| NH NHPI Alone | 0.29 | 0.30 | 0.25 |
| NH White Alone | 40.35 | 39.93 | 43.45 |
| NH AIAN & White | 1.41 | 1.75 | 0.16 |
| NH Asian & White | 1.02 | 0.87 | 0.71 |
| NH Black & White | 2.03 | 1.95 | 1.51 |
| NH AIAN and Black | 0.11 | 0.20 | 0.04 |
| NH Other Multi-Race | 0.88 | 0.72 | 0.57 |
| Hispanic | 31.18 | 33.16 | 31.89 |
| | Percent of Age 18+ | | |
| Non-Hispanic | 78.71 | 78.01 | 77.69 |
| NH AIAN Alone | 1.55 | 1.63 | 1.79 |
| NH Asian Alone | 4.64 | 4.34 | 3.94 |
| NH Black Alone | 15.06 | 14.47 | 15.67 |
| NH NHPI Alone | 0.24 | 0.25 | 0.21 |
| NH White Alone | 54.62 | 54.46 | 54.78 |
| NH AIAN & White | 1.21 | 1.64 | 0.16 |
| NH Asian & White | 0.39 | 0.31 | 0.32 |
| NH Black & White | 0.50 | 0.43 | 0.39 |
| NH AIAN and Black | 0.05 | 0.16 | 0.04 |
| NH Other Multi-Race | 0.45 | 0.31 | 0.41 |
| Hispanic | 21.28 | 21.98 | 22.31 |

Notes: The samples are people in housing units that had CAPI responses in the 2020 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The 2020 Census here means 2020 Census Edited File response data for the responding 2020 ACS housing units. The pre-disclosure avoidance ACS file is used in column 2. The weights are multiplied by five when calculating total population, since the weights are for 5-year rather than 1-year estimates. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.4D Population by Age and Race/Ethnicity, 2016-2019 ACS Housing Units with CAPI Responses

| | 2016 ACS, BW | 2016 AR, BW | 2017 ACS, BW | 2017 AR, BW | 2018 ACS, BW | 2018 AR, BW | 2019 ACS, BW | 2019 AR, BW |
|---|---|---|---|---|---|---|---|---|
| Total Population | 36,330,000 | 40,590,000 | 35,460,000 | 40,140,000 | 33,460,000 | 38,430,000 | 29,280,000 | 33,800,000 |
| | | | | | | | | |
| Age 0-17 | 28.35 | 31.17 | 27.58 | 30.52 | 26.72 | 29.44 | 26.16 | 28.85 |
| Age 18+ | 71.65 | 68.83 | 72.42 | 69.48 | 73.28 | 70.56 | 73.84 | 71.15 |
| | Age 0-17 | | | | | | | |
| Non-Hispanic | 64.71 | 63.68 | 64.78 | 63.67 | 64.15 | 63.06 | 64.20 | 63.69 |
| NH AIAN Alone | 2.11 | 2.11 | 2.13 | 2.16 | 2.14 | 2.17 | 2.10 | 2.22 |
| NH Asian Alone | 3.34 | 2.90 | 3.58 | 3.02 | 3.47 | 2.95 | 3.55 | 3.01 |
| NH Black Alone | 17.57 | 19.38 | 17.61 | 19.24 | 17.52 | 19.00 | 17.53 | 19.17 |
| NH NHPI Alone | 0.32 | 0.29 | 0.32 | 0.29 | 0.32 | 0.31 | 0.30 | 0.29 |
| NH White Alone | 37.42 | 35.67 | 37.16 | 35.66 | 36.76 | 35.35 | 36.88 | 35.77 |
| NH AIAN & White | 0.82 | 0.32 | 0.82 | 0.34 | 0.77 | 0.30 | 0.68 | 0.26 |
| NH Asian & White | 0.57 | 0.60 | 0.61 | 0.62 | 0.63 | 0.65 | 0.60 | 0.62 |
| NH Black & White | 1.85 | 1.74 | 1.83 | 1.66 | 1.80 | 1.65 | 1.90 | 1.69 |
| NH AIAN & Black | 0.16 | 0.08 | 0.15 | 0.08 | 0.15 | 0.08 | 0.14 | 0.07 |
| NH Other Multi-Race | 0.55 | 0.58 | 0.56 | 0.59 | 0.59 | 0.59 | 0.52 | 0.58 |
| Hispanic | 35.29 | 36.34 | 35.24 | 36.31 | 35.84 | 36.93 | 35.80 | 36.33 |
| | Age 18+ | | | | | | | |
| Non-Hispanic | 74.38 | 70.78 | 74.75 | 71.39 | 74.39 | 71.81 | 74.19 | 72.41 |
| NH AIAN Alone | 1.75 | 1.77 | 1.76 | 1.80 | 1.72 | 1.80 | 1.65 | 1.80 |
| NH Asian Alone | 4.77 | 4.64 | 4.87 | 4.77 | 4.75 | 4.60 | 4.79 | 4.62 |
| NH Black Alone | 16.37 | 17.51 | 16.61 | 17.75 | 16.69 | 17.84 | 16.74 | 18.14 |
| NH NHPI Alone | 0.26 | 0.25 | 0.25 | 0.25 | 0.28 | 0.26 | 0.26 | 0.25 |
| NH White Alone | 49.53 | 45.17 | 49.59 | 45.36 | 49.33 | 45.86 | 49.21 | 46.15 |
| NH AIAN & White | 0.69 | 0.30 | 0.66 | 0.31 | 0.61 | 0.28 | 0.53 | 0.25 |
| NH Asian & White | 0.25 | 0.27 | 0.25 | 0.29 | 0.24 | 0.28 | 0.24 | 0.29 |
| NH Black & White | 0.35 | 0.37 | 0.35 | 0.39 | 0.37 | 0.39 | 0.36 | 0.41 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NH AIAN & Black | 0.13 | 0.09 | 0.12 | 0.08 | 0.11 | 0.08 | 0.12 | 0.08 |
| NH Other Multi-Race | 0.29 | 0.41 | 0.29 | 0.41 | 0.29 | 0.41 | 0.27 | 0.41 |
| Hispanic | 25.62 | 29.22 | 25.25 | 28.61 | 25.61 | 28.19 | 25.81 | 27.59 |

Notes: The samples are people in housing units that have CAPI responses in the 2016-2019 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS files are used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.4E Population by Age and Race/Ethnicity, 2020 ACS Nonresponding Housing Units

| | 2020 Census, BW | ACS, TW-BW | ACS, EBW-BW | AR, BW |
|---|---|---|---|---|
| Total Population | 173,700,000 | 123,800,000 | 187,800,000 | 174,800,000 |
| | Percent by Age | | | |
| Age 0-17 | 24.26 | 22.81 | 23.77 | 26.46 |
| Age 18+ | 75.74 | 77.19 | 76.23 | 73.53 |
| | Percent of Age 0-17 | | | |
| Non-Hispanic | 71.01 | 70.27 | 71.51 | 70.56 |
| NH AIAN Alone | 0.71 | 0.51 | 0.44 | 0.96 |
| NH Asian Alone | 5.01 | 4.82 | 4.17 | 3.44 |
| NH Black Alone | 16.01 | 15.51 | 15.65 | 15.63 |
| NH NHPI Alone | 0.28 | 0.24 | 0.16 | 0.23 |
| NH White Alone | 43.48 | 43.48 | 45.89 | 47.13 |
| NH AIAN & White | 1.10 | 0.79 | 0.25 | 0.14 |
| NH Asian & White | 1.29 | 1.43 | 1.44 | 0.86 |
| NH Black & White | 2.18 | 2.47 | 2.85 | 1.55 |
| NH AIAN and Black | 0.12 | 0.14 | 0.05 | 0.04 |
| NH Other Multi-Race | 0.82 | 0.88 | 0.60 | 0.59 |
| Hispanic | 28.99 | 29.73 | 28.49 | 29.44 |
| | Percent of Age 18+ | | | |
| Non-Hispanic | 79.90 | 79.86 | 80.72 | 79.14 |
| NH AIAN Alone | 0.61 | 0.50 | 0.41 | 0.83 |
| NH Asian Alone | 5.99 | 5.88 | 5.39 | 5.12 |
| NH Black Alone | 14.19 | 14.08 | 14.14 | 14.37 |
| NH NHPI Alone | 0.21 | 0.21 | 0.17 | 0.19 |
| NH White Alone | 56.31 | 56.61 | 58.49 | 57.20 |
| NH AIAN & White | 1.05 | 0.91 | 0.22 | 0.14 |
| NH Asian & White | 0.50 | 0.51 | 0.61 | 0.40 |
| NH Black & White | 0.58 | 0.63 | 0.92 | 0.42 |
| NH AIAN and Black | 0.06 | 0.14 | 0.04 | 0.05 |
| NH Other Multi-Race | 0.44 | 0.40 | 0.33 | 0.43 |
| Hispanic | 20.11 | 20.17 | 19.29 | 20.85 |

Notes: The samples are people in housing units that are nonrespondents in the 2020 ACS. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The 2020 Census here means 2020 Census Edited File responses for nonresponding 2020 ACS housing units. The ACS TW-BW (EBW-BW) estimates are calculated by subtracting the ACS estimates using base weights from the final ACS estimates using TW (EBW). The weights are multiplied by five when calculating total population, since the weights are for 5-year rather than 1-year estimates. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.4F Percentage Point Difference Between 2020 ACS Nonresponding and Responding Housing Units by Age and Race/Ethnicity

| | 2020 Census, BW | ACS, TW | ACS, EBW | AR, BW |
|---|---|---|---|---|
| | Age Categories | | | |
| Age 0-17 | 3.94 | 2.43 | 3.39 | 4.62 |
| Age 18+ | -3.94 | -2.41 | -3.37 | -4.61 |
| | Age 0-17 | | | |

| | | | | |
|---|---|---|---|---|
| Non-Hispanic | -8.75 | -9.22 | -7.99 | -8.51 |
| NH AIAN Alone | -0.11 | -0.29 | -0.35 | -0.17 |
| NH Asian Alone | -1.41 | -1.46 | -2.11 | -1.23 |
| NH Black Alone | 6.64 | 6.70 | 6.84 | 5.92 |
| NH NHPI Alone | 0.07 | 0.05 | -0.03 | 0.05 |
| NH White Alone | -13.26 | -13.25 | -10.85 | -12.87 |
| NH AIAN & White | -0.20 | -0.68 | -1.22 | -0.01 |
| NH Asian & White | -0.81 | -0.80 | -0.79 | -0.36 |
| NH Black & White | 0.27 | 0.43 | 0.81 | 0.16 |
| NH AIAN and Black | 0.01 | 0.04 | -0.05 | 0.01 |
| NH Other Multi-Race | 0.05 | 0.03 | -0.25 | -0.01 |
| Hispanic | 8.75 | 9.22 | 7.99 | 8.51 |
| | Age 18+ | | | |
| Non-Hispanic | -7.45 | -7.24 | -6.37 | -7.45 |
| NH AIAN Alone | 0.00 | -0.10 | -0.18 | -0.02 |
| NH Asian Alone | -0.51 | -0.62 | -1.11 | -0.62 |
| NH Black Alone | 5.69 | 5.97 | 6.03 | 5.65 |
| NH NHPI Alone | 0.06 | 0.07 | 0.03 | 0.04 |
| NH White Alone | -12.79 | -12.46 | -10.59 | -12.61 |
| NH AIAN & White | -0.05 | -0.29 | -0.98 | 0.00 |
| NH Asian & White | -0.08 | -0.08 | 0.02 | -0.07 |
| NH Black & White | 0.16 | 0.19 | 0.47 | 0.11 |
| NH AIAN and Black | 0.01 | 0.05 | -0.05 | 0.02 |
| NH Other Multi-Race | 0.09 | 0.06 | -0.01 | 0.07 |
| Hispanic | 7.44 | 7.27 | 6.40 | 7.44 |

Notes: These are differences between the respective columns in Table 3.4E and Table 3.4A. Both ACS, TW and ACS, EBW use the ACS, BW column in Table 3.4A. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The results for nonresponding ACS housing units are displayed in Table 3.4E. Here the 2020 Census-BW and AR continue to use ACS housing unit base weights. To show what the ACS approaches predict for nonresponding housing units, we subtract the ACS estimates using the base weights from the ACS estimates using the respective final person weights (traditional weights vs. entropy balance weights). This means that the traditional and entropy balance population adjustments are attributed solely to the nonrespondents here.[77] The AR total population estimate is just 1.1 million higher than the 2020 Census-BW estimate here. The overall difference between AR and the 2020 Census-BW when combining responding and nonresponding housing units is 5.0 million. The traditional-weighted ACS shows higher shares of the voting-age population and voting-age non-Hispanic Whites and a lower share for non-Hispanic Blacks compared to the 2020 Census, but the shares are similar overall. The entropy-balance-weighted ACS shares are more different, with higher shares of the voting-age population and non-Hispanic Whites and lower shares of non-Hispanic Asians, non-Hispanic Blacks, and Hispanics.

[77] We did not have access to weights including all other adjustments besides the population adjustments. We thus do not try to evaluate the ACS against the 2020 Census regarding total population estimates for nonresponding housing units. Since the 2020 Census and AR estimates in Tables 3.4A-B use the same weights (and without population control adjustments), we can compare their total population estimates. We show below that population adjustments have little effect on the enhanced CVAP population shares by age and race/ethnicity.

When subtracting the shares in Tables 3.4A from those in 3.4E, Table 3.4F shows that the shares of non-Hispanic White and voting-age non-Hispanic Asian are smaller in nonresponding housing units, and the shares of non-Hispanic Blacks and Hispanics are higher. The magnitudes of the differences in distributions between responding and nonresponding housing units are very similar in the 2020 Census and AR estimates. The entropy-balance-weighted ACS shows a larger decrease in the share of non-Hispanic Asians, a smaller decrease in the share of non-Hispanic Whites, a larger increase in non-Hispanic Blacks, and a smaller increase in Hispanics compared to the 2020 Census and AR.

The distributions for ACS CAPI responses (Table 3.4C) and nonresponding housing units (Table 3.4E) are much more similar to each other than they are to those in all ACS responses (Table 3.4A). This makes sense, given that housing units with CAPI responses were randomly selected for follow-up, and nonresponding housing units are primarily units that were randomly not selected to be in CAPI follow-up.

Table 3.4G Percent of Population in 2020 ACS Nonresponding Housing Units by Age and Race/Ethnicity

| | 2020 Census, BW | ACS, TW | ACS, EBW | AR, BW |
|---|---|---|---|---|
| Total | 57.21 | 49.02 | 59.38 | 56.62 |
| | Age Categories | | | |
| Age 0-17 | 61.48 | 51.84 | 63.02 | 61.27 |
| Age 18+ | 55.96 | 48.26 | 58.32 | 55.13 |
| | Age 0-17 | | | |
| Non-Hispanic | 58.70 | 48.76 | 60.52 | 58.53 |
| NH AIAN Alone | 57.94 | 40.62 | 48.70 | 57.31 |
| NH Asian Alone | 55.48 | 45.24 | 53.08 | 53.82 |
| NH Black Alone | 73.16 | 65.45 | 75.16 | 71.81 |
| NH NHPI Alone | 67.60 | 58.02 | 59.12 | 67.33 |
| NH White Alone | 55.02 | 45.21 | 57.96 | 55.41 |
| NH AIAN & White | 57.50 | 36.56 | 22.70 | 59.44 |
| NH Asian & White | 49.45 | 40.83 | 52.37 | 52.61 |
| NH Black & White | 64.59 | 56.61 | 70.46 | 63.76 |
| NH AIAN and Black | 62.56 | 60.65 | 47.68 | 69.63 |
| NH Other Multi-Race | 62.97 | 52.70 | 54.60 | 60.84 |
| Hispanic | 69.57 | 60.95 | 70.31 | 68.99 |
| | Age 18+ | | | |
| Non-Hispanic | 53.76 | 46.09 | 56.47 | 52.90 |
| NH AIAN Alone | 56.07 | 43.75 | 49.18 | 54.60 |
| NH Asian Alone | 53.95 | 45.74 | 53.70 | 52.27 |
| NH Black Alone | 67.98 | 61.83 | 70.94 | 66.94 |
| NH NHPI Alone | 63.44 | 58.29 | 62.58 | 60.67 |
| NH White Alone | 50.88 | 43.32 | 54.23 | 50.17 |
| NH AIAN & White | 54.89 | 41.32 | 20.56 | 55.57 |
| NH Asian & White | 52.43 | 44.76 | 59.25 | 50.87 |
| NH Black & White | 63.53 | 56.89 | 74.11 | 62.29 |
| NH AIAN and Black | 61.32 | 60.14 | 39.30 | 65.87 |
| NH Other Multi-Race | 61.43 | 52.18 | 57.77 | 59.25 |
| Hispanic | 66.85 | 59.32 | 67.68 | 65.64 |

Notes: The samples are people in housing units that are in the initial 2020 ACS sample. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called

traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. 2020 Census response data come from the Census Edited File. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

In Table 3.4G we focus on how the proportions of each population group are divided between responding and nonresponding housing units. AR shares are quite close to those in the 2020 Census-BW. Though the characteristic distributions differ somewhat in levels between AR and the 2020 Census-BW, they both show similar changes between the responding and nonresponding housing units. The ACS approaches' nonresponding proportions are much less similar to the 2020 Census-BW.

In a robustness exercise not shown here, we produced results like those in Tables 3.4A, 3.4C, and 3.4E, but excluding housing units with AR enumeration, a proxy response, or whole household imputation in the 2020 Census. The proxies and whole household imputations likely produce more data containing greater error, and the AR enumerations are not a good comparator for AR, since they also use some of the same AR sources. The patterns are very similar to those in Tables 3.4 A, 3.4C, and 3.4E.

Table 3.5A Percent Citizens by Age and Race/Ethnicity, 2020 ACS Responding Housing Units

| | ACS, BW | AR, BW |
|---|---|---|
| Total | 94.79 | 93.91 |
| | Age Categories | |
| Age 0-17 | 97.80 | 98.61 |
| Age 18+ | 94.02 | 92.60 |
| | Age 0-17 | |
| Non-Hispanic | 98.27 | 98.72 |
| NH AIAN Alone | 99.88 | 99.10 |
| NH Asian Alone | 88.44 | 95.62 |
| NH Black Alone | 97.30 | 98.52 |
| NH NHPI Alone | 92.99 | 95.90 |
| NH White Alone | 99.37 | 99.04 |
| NH AIAN & White | 99.98 | 98.85 |
| NH Asian & White | 98.72 | 97.31 |
| NH Black & White | 99.59 | 98.96 |
| NH AIAN and Black | 99.72 | 97.95 |
| NH Other Multi-Race | 99.25 | 97.07 |
| Hispanic | 95.96 | 98.17 |
| | Age 18+ | |
| Non-Hispanic | 96.39 | 95.30 |
| NH AIAN Alone | 99.35 | 96.48 |
| NH Asian Alone | 72.96 | 75.28 |
| NH Black Alone | 95.90 | 95.85 |
| NH NHPI Alone | 82.89 | 82.42 |
| NH White Alone | 98.60 | 96.95 |
| NH AIAN & White | 99.82 | 97.44 |
| NH Asian & White | 95.19 | 91.90 |

| | | |
|---|---|---|
| NH Black & White | 96.78 | 97.15 |
| NH AIAN and Black | 99.18 | 96.03 |
| NH Other Multi-Race | 96.56 | 87.21 |
| Hispanic | 77.92 | 75.09 |

Notes: The samples are people in housing units that are responding units in the 2020 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS file is used in column 1 here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.5B Percent Citizens by Age and Race/Ethnicity, 2016-2019 ACS Responding Housing Units

| | 2016 ACS, BW | 2016 AR, BW | 2017 ACS, BW | 2017 AR, BW | 2018 ACS, BW | 2018 AR, BW | 2019 ACS, BW | 2019 AR, BW |
|---|---|---|---|---|---|---|---|---|
| Total Population | 94.55 | 93.31 | 94.50 | 93.27 | 94.73 | 93.53 | 94.98 | 94.05 |
| | | | | | | | | |
| Age 0-17 | 97.82 | 98.47 | 97.73 | 98.44 | 97.83 | 98.44 | 97.88 | 98.63 |
| Age 18+ | 93.65 | 91.72 | 93.61 | 91.67 | 93.89 | 92.08 | 94.22 | 92.74 |
| | Age 0-17 | | | | | | | |
| Non-Hispanic | 98.34 | 98.70 | 98.20 | 98.67 | 98.29 | 98.67 | 98.29 | 98.81 |
| NH AIAN Alone | 99.86 | 99.24 | 99.62 | 99.27 | 99.87 | 99.45 | 99.90 | 99.45 |
| NH Asian Alone | 87.57 | 92.09 | 87.23 | 92.44 | 87.96 | 92.56 | 88.23 | 93.20 |
| NH Black Alone | 98.02 | 98.14 | 97.74 | 98.13 | 97.81 | 97.97 | 97.79 | 98.28 |
| NH NHPI Alone | 93.89 | 93.54 | 92.95 | 93.65 | 92.79 | 93.98 | 92.20 | 92.42 |
| NH White Alone | 99.32 | 99.37 | 99.27 | 99.32 | 99.29 | 99.33 | 99.30 | 99.43 |
| NH AIAN & White | 99.94 | 99.67 | 99.94 | 99.50 | 99.89 | 99.70 | 99.96 | 99.54 |
| NH Asian & White | 98.11 | 97.92 | 97.93 | 97.93 | 97.70 | 97.94 | 97.41 | 97.81 |
| NH Black & White | 99.76 | 99.42 | 99.85 | 99.47 | 99.79 | 99.51 | 99.78 | 99.51 |
| NH AIAN & Black | 99.71 | 98.96 | 100.00 | 99.37 | 99.74 | 98.75 | 100.00 | 99.53 |
| NH Other Multi-Race | 99.21 | 98.03 | 99.35 | 98.25 | 99.13 | 98.32 | 98.71 | 97.99 |
| Hispanic | 95.70 | 97.65 | 95.85 | 97.61 | 95.98 | 97.61 | 96.10 | 97.98 |
| | Age 18+ | | | | | | | |
| Non-Hispanic | 96.29 | 95.53 | 96.15 | 95.39 | 96.31 | 95.51 | 96.41 | 95.68 |
| NH AIAN Alone | 99.43 | 97.21 | 99.36 | 96.96 | 99.44 | 97.13 | 99.51 | 97.41 |
| NH Asian Alone | 70.65 | 69.67 | 70.24 | 69.82 | 71.51 | 70.53 | 72.49 | 71.49 |
| NH Black Alone | 95.83 | 95.38 | 95.77 | 95.36 | 95.82 | 95.33 | 95.92 | 95.55 |
| NH NHPI Alone | 84.89 | 83.28 | 86.29 | 83.35 | 85.39 | 83.05 | 84.32 | 83.96 |
| NH White Alone | 98.42 | 97.72 | 98.37 | 97.63 | 98.44 | 97.71 | 98.52 | 97.84 |
| NH AIAN & White | 99.72 | 98.36 | 99.72 | 98.25 | 99.71 | 98.44 | 99.66 | 98.08 |
| NH Asian & White | 91.45 | 91.12 | 92.00 | 91.25 | 92.50 | 91.27 | 91.96 | 91.16 |
| NH Black & White | 98.09 | 96.80 | 97.90 | 96.76 | 98.06 | 96.91 | 98.20 | 97.43 |
| NH AIAN & Black | 99.40 | 97.49 | 99.26 | 97.57 | 99.26 | 97.60 | 99.41 | 97.38 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NH Other Multi-Race | 95.04 | 86.47 | 95.87 | 87.14 | 95.83 | 87.35 | 95.84 | 88.13 |
| Hispanic | 74.22 | 68.31 | 75.22 | 69.13 | 76.46 | 71.03 | 77.99 | 73.84 |

Notes: The samples are people in housing units that are responding units in the 2016-2019 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS is used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.5C Percent Citizens by Age and Race/Ethnicity, 2020 ACS Housing Units with CAPI Responses

| | ACS, BW | AR, BW |
|---|---|---|
| Total | 91.25 | 90.45 |
| | Age Categories | |
| Age 0-17 | 96.85 | 98.10 |
| Age 18+ | 89.57 | 87.77 |
| | Age 0-17 | |
| Non-Hispanic | 98.23 | 98.24 |
| NH AIAN Alone | 99.93 | 99.44 |
| NH Asian Alone | 86.97 | 92.84 |
| NH Black Alone | 97.60 | 98.49 |
| NH NHPI Alone | 92.56 | 94.43 |
| NH White Alone | 99.33 | 98.49 |
| NH AIAN & White | 100.0 | 98.73 |
| NH Asian & White | 95.39 | 95.82 |
| NH Black & White | 99.85 | 98.43 |
| NH AIAN and Black | 99.90 | 97.00 |
| NH Other Multi-Race | 98.34 | 96.13 |
| Hispanic | 94.08 | 97.80 |
| | Age 18+ | |
| Non-Hispanic | 96.09 | 93.94 |
| NH AIAN Alone | 99.65 | 97.56 |
| NH Asian Alone | 67.87 | 67.97 |
| NH Black Alone | 95.37 | 95.13 |
| NH NHPI Alone | 78.43 | 73.62 |
| NH White Alone | 98.43 | 95.51 |
| NH AIAN & White | 99.86 | 97.34 |
| NH Asian & White | 87.51 | 88.15 |
| NH Black & White | 97.13 | 96.38 |
| NH AIAN and Black | 99.32 | 95.61 |
| NH Other Multi-Race | 94.08 | 83.50 |
| Hispanic | 66.43 | 66.30 |

Notes: The samples are people in housing units that are responding units in the 2020 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS file is used in column 1 here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.5D Percent Citizens by Age and Race/Ethnicity, 2016-2019 ACS Housing Units with CAPI Responses

| | 2016 ACS, BW | 2016 AR, BW | 2017 ACS, BW | 2017 AR, BW | 2018 ACS, BW | 2018 AR, BW | 2019 ACS, BW | 2019 AR, BW |
|---|---|---|---|---|---|---|---|---|
| Total Population | 88.75 | 86.78 | 89.08 | 87.11 | 89.25 | 87.64 | 89.39 | 88.47 |
| | | | | | | | | |
| Age 0-17 | 96.55 | 97.77 | 96.42 | 97.68 | 96.51 | 97.70 | 96.38 | 97.87 |
| Age 18+ | 85.66 | 81.82 | 86.29 | 82.46 | 86.60 | 83.44 | 86.92 | 84.66 |
| | Age 0-17 | | | | | | | |
| Non-Hispanic | 97.85 | 98.05 | 97.56 | 97.96 | 97.79 | 97.96 | 97.76 | 98.11 |
| NH AIAN Alone | 99.88 | 99.35 | 99.85 | 99.41 | 99.93 | 99.47 | 99.93 | 99.61 |
| NH Asian Alone | 82.95 | 87.92 | 82.42 | 88.25 | 83.45 | 88.41 | 84.79 | 90.49 |
| NH Black Alone | 98.02 | 98.15 | 97.56 | 98.10 | 97.93 | 98.09 | 97.75 | 98.24 |
| NH NHPI Alone | 93.09 | 91.33 | 89.41 | 89.85 | 90.11 | 91.79 | 88.76 | 87.93 |
| NH White Alone | 98.94 | 98.82 | 98.87 | 98.69 | 98.94 | 98.68 | 98.92 | 98.73 |
| NH AIAN & White | 99.92 | 99.50 | 100.00 | 99.00 | 99.89 | 99.60 | 100.00 | 99.44 |
| NH Asian & White | 93.31 | 94.58 | 91.59 | 94.08 | 92.40 | 93.10 | 89.24 | 92.31 |
| NH Black & White | 99.82 | 99.22 | 99.62 | 99.17 | 99.79 | 99.13 | 99.89 | 99.10 |
| NH AIAN & Black | 100.00 | 98.22 | 100.00 | 98.89 | 100.00 | 98.02 | 100.00 | 99.44 |
| NH Other Multi-Race | 98.28 | 96.80 | 99.09 | 97.32 | 98.87 | 97.08 | 99.06 | 96.71 |
| Hispanic | 94.17 | 97.28 | 94.32 | 97.17 | 94.23 | 97.26 | 93.91 | 97.46 |
| | Age 18+ | | | | | | | |
| Non-Hispanic | 94.39 | 92.58 | 94.45 | 92.60 | 94.73 | 92.94 | 94.89 | 93.18 |
| NH AIAN Alone | 99.65 | 97.41 | 99.71 | 97.12 | 99.70 | 97.48 | 99.71 | 97.48 |
| NH Asian Alone | 60.66 | 60.01 | 60.53 | 60.72 | 62.68 | 61.81 | 64.12 | 63.01 |
| NH Black Alone | 94.45 | 93.97 | 94.74 | 94.27 | 95.04 | 94.52 | 94.99 | 94.63 |
| NH NHPI Alone | 77.61 | 74.17 | 77.75 | 71.95 | 79.59 | 74.24 | 75.69 | 71.57 |
| NH White Alone | 97.49 | 95.37 | 97.56 | 95.30 | 97.61 | 95.47 | 97.80 | 95.67 |
| NH AIAN & White | 99.74 | 97.98 | 99.86 | 98.00 | 99.87 | 97.83 | 99.68 | 97.72 |
| NH Asian & White | 77.35 | 83.40 | 79.58 | 83.31 | 78.58 | 81.06 | 77.03 | 80.73 |
| NH Black & White | 97.66 | 96.06 | 96.80 | 96.37 | 97.64 | 96.22 | 98.10 | 97.52 |
| NH AIAN & Black | 99.93 | 96.26 | 99.75 | 96.66 | 99.92 | 97.07 | 99.91 | 97.82 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NH Other Multi-Race | 94.14 | 82.15 | 94.45 | 82.41 | 93.13 | 82.85 | 94.59 | 84.17 |
| Hispanic | 60.32 | 55.76 | 62.14 | 57.19 | 63.01 | 59.27 | 64.00 | 62.29 |

Notes: The samples are people in housing units that are responding units in the 2016-2019 ACS. BW means ACS housing unit base weights. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS files are used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.5E Percent Citizens by Age and Race/Ethnicity, 2020 ACS Nonresponding Housing Units

| | ACS, TW-BW | ACS, EBW-BW | AR, BW |
|---|---|---|---|
| Total | 93.11 | 92.58 | 91.22 |
| | Age Categories | | |
| Age 0-17 | 97.39 | 97.33 | 98.33 |
| Age 18+ | 91.96 | 91.09 | 88.67 |
| | Age 0-17 | | |
| Non-Hispanic | 98.18 | 98.17 | 98.42 |
| NH AIAN Alone | 99.81 | 99.86 | 98.79 |
| NH Asian Alone | 87.72 | 86.88 | 94.44 |
| NH Black Alone | 97.62 | 97.70 | 98.78 |
| NH NHPI Alone | 92.59 | 92.45 | 94.29 |
| NH White Alone | 99.43 | 99.27 | 98.65 |
| NH AIAN & White | 100.0 | 100.0 | 98.59 |
| NH Asian & White | 98.01 | 97.47 | 96.27 |
| NH Black & White | 99.60 | 99.59 | 98.80 |
| NH AIAN and Black | 99.93 | 99.79 | 98.65 |
| NH Other Multi-Race | 98.85 | 98.93 | 96.94 |
| Hispanic | 95.51 | 95.25 | 98.12 |
| | Age 18+ | | |
| Non-Hispanic | 96.14 | 95.48 | 93.79 |
| NH AIAN Alone | 99.28 | 99.20 | 94.91 |
| NH Asian Alone | 70.11 | 68.57 | 71.17 |
| NH Black Alone | 95.16 | 95.03 | 95.46 |
| NH NHPI Alone | 81.24 | 81.18 | 79.56 |
| NH White Alone | 98.41 | 98.09 | 95.50 |
| NH AIAN & White | 99.84 | 99.69 | 96.82 |
| NH Asian & White | 93.78 | 92.60 | 88.79 |
| NH Black & White | 99.83 | 96.93 | 96.99 |
| NH AIAN and Black | 98.86 | 98.43 | 96.18 |
| NH Other Multi-Race | 95.33 | 94.59 | 84.86 |
| Hispanic | 73.43 | 72.81 | 69.20 |

Notes: The samples are people in housing units that are nonrespondents in the 2020 ACS. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The ACS TW-BW (EBW-BW) estimates are calculated by subtracting the ACS estimates using base weights from the final ACS estimates using TW (EBW). The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.5F Percent Citizens by Age and Race/Ethnicity, 2020 ACS Responding and Nonresponding Housing Units

| | ACS, TW | ACS, EBW | AR, BW |
|---|---|---|---|
| Total | 93.94 | 93.48 | 92.39 |
| | Age Categories | | |
| Age 0-17 | 97.59 | 97.51 | 98.44 |
| Age 18+ | 92.98 | 92.32 | 90.43 |
| | Age 0-17 | | |
| Non-Hispanic | 98.24 | 98.22 | 98.55 |
| NH AIAN Alone | 99.85 | 99.87 | 98.92 |
| NH Asian Alone | 88.10 | 87.60 | 94.98 |
| NH Black Alone | 97.50 | 97.60 | 98.71 |
| NH NHPI Alone | 92.80 | 92.71 | 94.82 |
| NH White Alone | 99.45 | 99.36 | 98.82 |
| NH AIAN & White | 100.0 | 99.99 | 98.69 |
| NH Asian & White | 98.41 | 98.05 | 96.76 |
| NH Black & White | 99.60 | 99.59 | 98.86 |
| NH AIAN and Black | 99.84 | 99.75 | 98.44 |
| NH Other Multi-Race | 99.05 | 99.09 | 96.99 |
| Hispanic | 95.69 | 95.46 | 98.13 |
| | Age 18+ | | |
| Non-Hispanic | 96.32 | 95.94 | 94.50 |
| NH AIAN Alone | 99.34 | 99.30 | 95.62 |
| NH Asian Alone | 71.58 | 70.52 | 73.13 |
| NH Black Alone | 95.45 | 95.30 | 95.58 |
| NH NHPI Alone | 81.99 | 81.96 | 80.68 |
| NH White Alone | 98.74 | 98.45 | 96.22 |
| NH AIAN & White | 99.83 | 99.80 | 97.10 |
| NH Asian & White | 94.54 | 93.65 | 90.32 |
| NH Black & White | 96.94 | 96.79 | 97.05 |
| NH AIAN and Black | 98.99 | 98.89 | 96.13 |
| NH Other Multi-Race | 95.95 | 95.38 | 85.81 |
| Hispanic | 75.26 | 74.49 | 71.23 |

Notes: The samples are people in housing units that are in the initial 2020 ACS sample. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. AR here means that the household rosters come from administrative record sources, and their age, race/ethnicity, and citizenship are from numerous survey and administrative record sources and modeling. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Tables 3.5A-F display citizen shares by age and race/ethnicity for the 2020 and 2016-2019 samples like those used for Tables 3.4A-E, though without the 2020 Census, which does not contain citizenship. Entropy balance weights lower the overall estimated citizen share in 2020 by 0.46 percentage points compared to traditional weights (Table 3.5F). In responding 2020 ACS housing units, the AR and modeling citizen share is 0.88 percentage points lower than when using the traditional approaches (Table 3.5A). The AR/modeling share in 2020 ACS responding housing units is higher for children (0.81 percentage points), but 1.42 percentage points lower for the voting-age population. The ACS

approaches have lower citizen shares for non-Hispanic Asians and Hispanic children and higher ones for non-Hispanic Whites and voting-age Hispanics. These patterns are present in nonresponding ACS housing units as well. The differences across methods are generally larger for nonresponding than responding units.

ACS-AR citizen share differences in responding housing units narrowed significantly over time (2.24, 1.23, 1.20, 0.93, and 0.88 percentage points higher in the ACS in 2016, 2017, 2018, 2019, and 2020, respectively). This is especially true for voting-age Hispanics (5.91, 6.09, 5.43, 4.15, and 2.83 percentage points higher in the ACS in 2016, 2017, 2018, 2019, and 2020, respectively). ACS-AR citizen share differences are smaller in housing units with CAPI responses.

Bringing the age-race/ethnicity and citizenship results together, Table 3.4A shows a higher AR total population and voting-age Hispanic share estimate in responding housing units compared to the ACS, and Table 3.5A displays a lower AR voting-age Hispanic citizen share for those units. Of the additional voting-age Hispanics in AR compared to the ACS in responding housing units, 78.2 percent are noncitizens. Similarly, Brown et al. (2023) suggest that 84.2 percent of the additional people in AR compared to the 2020 PEP estimates are noncitizens.

## *3.3 Comparing Methods in Responding and Nonresponding Housing Units in 2016-2020*

In Table 3.6 we repeat the exercises in Tables 3.4 and 3.5, now pooling all years from 2016-2020. AR has a higher population; higher shares of children, non-Hispanic AIAN, non-Hispanic Black, and Hispanic; and a lower share of non-Hispanic Asians in responding ACS housing units. Unlike in 2020 alone, AR has a lower share for non-Hispanic Whites. For nonresponding ACS housing units, the AR population is lower; AR shares are higher for children, non-Hispanic AIAN, Hispanics, and non-Hispanic White children; and they are lower for non-Hispanic Asians and non-Hispanic Blacks.

Table 3.6 Percent of Population by Age and Race/Ethnicity, Responding vs. Nonresponding 2016-2020 ACS Housing Units

| | ACS in Responding Units, BW | AR in Responding Units, BW | ACS, TW-BW | ACS, EBW-BW | AR in Nonresponding Units, BW |
|---|---|---|---|---|---|
| Total Population | 183,200,000 | 190,500,000 | 135,300,000 | 135,300,000 | 117,300,000 |
| | Percent by Age | | | | |
| Age 0-17 | 21.09 | 22.90 | 25.48 | 25.45 | 28.70 |
| Age 18+ | 78.91 | 77.10 | 74.52 | 74.55 | 71.30 |
| | Percent of Age 0-17 | | | | |
| Non-Hispanic | 80.28 | 78.18 | 68.66 | 68.64 | 68.18 |
| NH AIAN Alone | 0.82 | 1.11 | 0.62 | 0.62 | 0.91 |
| NH Asian Alone | 5.79 | 4.90 | 4.18 | 4.15 | 3.24 |
| NH Black Alone | 9.18 | 10.52 | 18.21 | 18.13 | 17.56 |
| NH NHPI Alone | 0.16 | 0.17 | 0.22 | 0.22 | 0.25 |
| NH White Alone | 59.09 | 57.20 | 40.57 | 40.59 | 43.05 |
| NH AIAN & White | 0.70 | 0.31 | 0.40 | 0.37 | 0.14 |
| NH Asian & White | 1.86 | 1.54 | 1.13 | 1.18 | 0.78 |

| | | | | | |
|---|---|---|---|---|---|
| NH Black & White | 1.90 | 1.68 | 2.45 | 2.59 | 1.60 |
| NH AIAN and Black | 0.08 | 0.05 | 0.12 | 0.10 | 0.05 |
| NH Other Multi-Race | 0.69 | 0.70 | 0.74 | 0.70 | 0.62 |
| Hispanic | 19.72 | 21.82 | 31.34 | 31.36 | 31.82 |
| | Percent of Age 18+ | | | | |
| Non-Hispanic | 87.81 | 86.15 | 77.95 | 77.88 | 76.21 |
| NH AIAN Alone | 0.59 | 0.87 | 0.60 | 0.63 | 0.81 |
| NH Asian Alone | 6.05 | 5.79 | 5.57 | 5.53 | 5.00 |
| NH Black Alone | 8.32 | 9.05 | 16.71 | 16.64 | 16.18 |
| NH NHPI Alone | 0.13 | 0.15 | 0.21 | 0.21 | 0.22 |
| NH White Alone | 71.02 | 68.89 | 52.95 | 52.89 | 52.55 |
| NH AIAN & White | 0.54 | 0.24 | 0.43 | 0.39 | 0.14 |
| NH Asian & White | 0.48 | 0.46 | 0.42 | 0.46 | 0.37 |
| NH Black & White | 0.32 | 0.31 | 0.57 | 0.67 | 0.42 |
| NH AIAN and Black | 0.06 | 0.04 | 0.12 | 0.09 | 0.05 |
| NH Other Multi-Race | 0.29 | 0.37 | 0.37 | 0.37 | 0.47 |
| Hispanic | 12.16 | 13.83 | 22.08 | 22.06 | 23.74 |

Notes: The samples are people in housing units in the initial 2016-2020 ACS sample. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The ACS TW-BW (EBW-BW) estimates are calculated by subtracting the ACS estimates using base weights from the final ACS estimates using TW (EBW). The pre-disclosure avoidance ACS file is used in column 1 here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The responding ACS housing unit citizen shares in Table 3.7A include not only estimates from the ACS and AR/modeling, but also ones using the ACS roster combined with AR/modeling when the ACS citizenship value is an edit or an imputation (as in enhanced CVAP) and using the ACS roster combined with AR/modeling for all citizenship values. Table 3.7B shows the estimates for nonresponding housing units. The decomposition of the difference in estimates for the nonresponding housing units in AR vs. the ACS with entropy balance weights incorporated into the processing of the 2020 data uses the estimates in Tables 3.7A-B and is shown in Table 3.7C. It allows us to see how each change affects the citizen share estimates. Replacing ACS edits and imputations with AR/modeling reduces the overall citizen share estimate by 0.16 percentage points, and most of the differences by age and race/ethnicity are smaller than that. Inserting AR/modeling in place of as-reported ACS citizenship reduces the citizen share estimate another 0.04 percentage points. Using AR household rosters instead of ACS rosters when populating nonresponding housing units lowers the citizen share estimate by 0.91 percentage points. Finally, the difference in the estimated unit nonresponse bias when populating nonresponding housing units with person data rather than applying probability of selection and nonresponse adjustments to the responding person weights reduces the citizen share estimate by 0.77 percentage points.

Table 3.7A Percent Citizens by Age and Race/Ethnicity, Responding 2016-2020 ACS Housing Units

| | ACS Roster and Citizenship, BW | ACS Roster, AR and Modeling in Place of ACS Edits and Imputes in Responding Housing Units, BW | ACS Roster, AR and Modeling in Place of All ACS Citizenship Values | AR Rosters, AR and Modeling, BW |
|---|---|---|---|---|
| Total | 94.70 | 94.54 | 94.50 | 93.59 |
| | Age Categories | | | |
| Age 0-17 | 97.81 | 97.70 | 97.46 | 98.51 |
| Age 18+ | 93.87 | 93.69 | 93.70 | 92.12 |
| | Age 0-17 | | | |
| Non-Hispanic | 98.28 | 98.19 | 98.24 | 98.71 |
| NH AIAN Alone | 99.82 | 99.79 | 99.51 | 99.32 |
| NH Asian Alone | 87.85 | 87.53 | 88.31 | 92.96 |
| NH Black Alone | 97.77 | 97.60 | 97.39 | 98.18 |
| NH NHPI Alone | 92.99 | 92.85 | 92.87 | 93.76 |
| NH White Alone | 99.30 | 99.24 | 99.29 | 99.32 |
| NH AIAN & White | 99.95 | 99.94 | 99.78 | 99.55 |
| NH Asian & White | 97.93 | 97.83 | 97.79 | 97.84 |
| NH Black & White | 99.76 | 99.73 | 99.58 | 99.42 |
| NH AIAN and Black | 99.83 | 99.79 | 99.54 | 99.04 |
| NH Other Multi-Race | 99.14 | 99.11 | 98.76 | 98.02 |
| Hispanic | 95.91 | 95.70 | 94.28 | 97.76 |
| | Age 18+ | | | |
| Non-Hispanic | 96.31 | 96.23 | 96.25 | 95.49 |
| NH AIAN Alone | 99.42 | 99.45 | 99.15 | 97.08 |
| NH Asian Alone | 71.50 | 71.21 | 72.26 | 71.07 |
| NH Black Alone | 95.84 | 95.67 | 95.59 | 95.47 |
| NH NHPI Alone | 84.87 | 84.16 | 84.87 | 83.25 |
| NH White Alone | 98.46 | 98.41 | 98.36 | 97.61 |
| NH AIAN & White | 99.74 | 99.77 | 99.58 | 98.22 |
| NH Asian & White | 92.55 | 92.51 | 92.45 | 91.30 |
| NH Black & White | 97.81 | 97.79 | 97.59 | 97.01 |
| NH AIAN and Black | 99.30 | 99.42 | 99.15 | 97.37 |
| NH Other Multi-Race | 95.80 | 95.74 | 95.26 | 87.26 |
| Hispanic | 76.25 | 75.36 | 75.35 | 71.14 |

Notes: The samples are people in housing units that responded in the 2016-2020 ACS. BW means ACS housing unit base weights. AR rosters here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The pre-disclosure avoidance ACS file is used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.7B Percent Citizens by Age and Race/Ethnicity, Nonresponding 2016-2020 ACS Housing Units

| | ACS, TW-BW | ACS, EBW-BW | Adjusted AR in Nonresponding Housing Units, BW | AR in Nonresponding Housing Units, BW |
|---|---|---|---|---|
| Total | 91.49 | 91.48 | 90.55 | 89.60 |
| | Age Categories | | | |
| Age 0-17 | 97.18 | 97.16 | 97.31 | 98.12 |
| Age 18+ | 89.55 | 89.54 | 87.73 | 86.17 |
| | Age 0-17 | | | |
| Non-Hispanic | 98.05 | 98.04 | 97.73 | 98.26 |
| NH AIAN Alone | 99.77 | 99.72 | 99.05 | 98.58 |
| NH Asian Alone | 86.38 | 86.34 | 87.55 | 92.98 |
| NH Black Alone | 97.98 | 97.97 | 98.00 | 98.58 |
| NH NHPI Alone | 93.22 | 92.60 | 93.14 | 94.05 |
| NH White Alone | 99.19 | 99.17 | 98.53 | 98.60 |
| NH AIAN & White | 99.96 | 99.97 | 98.84 | 98.44 |
| NH Asian & White | 96.44 | 96.44 | 95.53 | 95.54 |
| NH Black & White | 99.72 | 99.71 | 99.07 | 98.76 |
| NH AIAN and Black | 99.99 | 99.95 | 99.40 | 98.65 |
| NH Other Multi-Race | 98.88 | 98.88 | 97.63 | 96.54 |
| Hispanic | 95.28 | 95.22 | 95.77 | 97.83 |
| | Age 18+ | | | |
| Non-Hispanic | 95.10 | 95.05 | 93.83 | 93.09 |
| NH AIAN Alone | 99.32 | 99.20 | 96.82 | 94.45 |
| NH Asian Alone | 67.03 | 66.84 | 68.05 | 67.91 |
| NH Black Alone | 95.12 | 95.07 | 95.29 | 95.09 |
| NH NHPI Alone | 81.65 | 81.02 | 78.28 | 77.37 |
| NH White Alone | 98.02 | 97.97 | 95.80 | 94.99 |
| NH AIAN & White | 99.74 | 99.66 | 98.20 | 96.65 |
| NH Asian & White | 89.65 | 89.96 | 88.75 | 87.54 |
| NH Black & White | 97.74 | 97.53 | 97.51 | 96.74 |
| NH AIAN and Black | 99.37 | 99.44 | 97.96 | 95.91 |
| NH Other Multi-Race | 95.06 | 94.76 | 91.94 | 83.46 |
| Hispanic | 69.96 | 70.11 | 68.14 | 63.92 |

Notes: The samples are people in housing units in the initial 2016-2020 ACS sample. BW means ACS housing unit base weights. TW means final person weights using the ACS production method used prior to 2020 (called traditional here). EBW is the final 2016-2020 ACS version of the person weights, which incorporates entropy balance weights in the processing of the 2020 data. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The ACS TW-BW (EBW-BW) estimates are calculated by subtracting the ACS estimates using base weights from the final ACS estimates using TW (EBW). The third column of Table 3.8B displays estimates from column 4 plus the percentage point difference between column 2 and column 4 in Table 3.8A. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.7C Decomposition of Difference in Percent Citizens between ACS, EBW-BW and AR in Nonresponding 2016-2020 ACS Housing Units

| | AR and Modeling in Place of ACS Edits and Imputes in Responding Housing Units | AR and Modeling in Place of As-Reported ACS Citizenship Value | AR Rosters in Place of ACS Rosters | Difference in AR vs. ACS EBW Change for Nonresponding vs. Responding Housing Units | Total Difference in Percent Citizens between ACS, EBW-BW and AR |
|---|---|---|---|---|---|
| Total | -0.16 | -0.04 | -0.91 | -0.77 | -1.88 |
| | Age Categories | | | | |
| Age 0-17 | -0.11 | -0.24 | 1.05 | 0.27 | 0.97 |
| Age 18+ | -0.18 | 0.01 | -1.58 | -1.63 | -3.38 |
| | Age 0-17 | | | | |
| Non-Hispanic | -0.09 | 0.05 | 0.47 | -0.21 | 0.22 |
| NH AIAN Alone | -0.03 | -0.28 | -0.19 | -0.65 | -1.15 |
| NH Asian Alone | -0.32 | 0.78 | 4.64 | 1.53 | 6.64 |
| NH Black Alone | -0.18 | -0.21 | 0.79 | 0.21 | 0.61 |
| NH NHPI Alone | -0.14 | 0.02 | 0.89 | 0.68 | 1.45 |
| NH White Alone | -0.06 | 0.05 | 0.03 | -0.59 | -0.57 |
| NH AIAN & White | 0.00 | -0.17 | -0.23 | -1.13 | -1.53 |
| NH Asian & White | -0.10 | -0.05 | 0.05 | -0.81 | -0.90 |
| NH Black & White | -0.03 | -0.15 | -0.16 | -0.61 | -0.95 |
| NH AIAN and Black | -0.04 | -0.25 | -0.50 | -0.51 | -1.31 |
| NH Other Multi-Race | -0.02 | -0.35 | -0.74 | -1.22 | -2.34 |
| Hispanic | -0.21 | -1.42 | 3.48 | 0.75 | 2.60 |
| | Age 18+ | | | | |
| Non-Hispanic | -0.08 | 0.02 | -0.76 | -1.14 | -1.96 |
| NH AIAN Alone | 0.02 | -0.30 | -2.07 | -2.40 | -4.74 |
| NH Asian Alone | -0.29 | 1.05 | -1.19 | 1.49 | 1.07 |
| NH Black Alone | -0.17 | -0.08 | -0.12 | 0.40 | 0.02 |
| NH NHPI Alone | -0.71 | 0.71 | -1.62 | -2.04 | -3.66 |
| NH White Alone | -0.05 | -0.05 | -0.75 | -2.12 | -2.98 |
| NH AIAN & White | 0.03 | -0.20 | -1.35 | -1.49 | -3.01 |
| NH Asian & White | -0.04 | -0.06 | -1.15 | -1.17 | -2.41 |
| NH Black & White | -0.02 | -0.19 | -0.58 | 0.01 | -0.80 |
| NH AIAN and Black | 0.12 | -0.27 | -1.77 | -1.60 | -3.53 |
| NH Other Multi-Race | -0.06 | -0.48 | -8.00 | -2.76 | -11.30 |
| Hispanic | -0.89 | -0.01 | -4.21 | -1.08 | -6.19 |

Notes: The first column is the second column of Table 3.7A minus the first column of Table 3.7A. The second column is the third column of Table 3.7A minus the second column of Table 3.7A. The third column is the fourth column of Table 3.7A minus the third column of Table 3.7A. The fourth column is (the fourth column of Table 3.7B minus the fourth column of Table 3.7A) minus (the second column of Table 3.7B minus the first column of Table 3.7A). The fifth column is the fourth column of Table 3.7B minus the second column of Table 3.7B. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The changes when switching from ACS to AR rosters are quite heterogeneous. The citizen share rises by 1.05 percentage points among all children and 3.48 percentage points for Hispanic children, while it drops 8.0 percentage points for voting-age non-Hispanic other multi-race. The difference in estimated

unit nonresponse bias also varies significantly, where it raises the citizen share estimate by about 1.5 percentage points for voting-age non-Hispanic Asians and lowers it by 2.12 percentage points for voting-age non-Hispanic Whites.

The AR citizen share estimate for voting-age Hispanics in nonresponding housing units is 6.19 percentage points lower than when using the ACS that incorporates entropy balance weights into the processing of the 2020 data. The largest contributor is the change in household rosters (4.21 percentage points), followed by the difference in unit nonresponse bias (1.08 percentage points) and the switch from ACS edits and imputations to AR and modeling (0.89 percentage points).

The third column of Table 3.7B displays estimates from AR in nonresponding ACS housing units, adjusted by the difference between column 2 (the ACS roster with AR and modeling in place of ACS citizenship edits and imputations, which is used in enhanced CVAP) and column 4 (AR rosters and AR/modeled citizenship) in Table 3.7A. If one were to assume that enhanced CVAP is the best estimate of the citizen share in responding ACS housing units, and the change between the responding and nonresponding ACS housing unit AR estimates is the best measure of the difference between responding and nonresponding units, then this adjusted measure would be a good measure of the citizen share in nonresponding units.[78] Overall, the distances between each measure and the adjusted measure are about the same. The ACS measures are closer for children. AR is closer for voting-age non-Hispanic people. The category with the largest difference in differences is voting-age Hispanics, where the ACS estimates are 1.82 to 1.97 percentage points away from the adjusted measure, vs. 4.22 for AR. Comprehending the ACS vs. AR differences in the citizenship composition of household rosters for voting-age Hispanics in responding ACS housing units is thus key to understanding the overall differences between methods.

A comparison of the total population estimates in Table 3.6 shows that the AR estimate for responding and nonresponding housing units combined using base weights is 10.7 million below the overall ACS estimate that is controlled to the PEP estimate. AR may thus be undercounting the population, at least when restricted to the ACS sample.[79] Since the AR total for responding ACS housing units is 7.3 million higher than that of the ACS, the ACS appears to be undercounting the population even more than AR. In a context where the sources may be undercounting the population, it is instructive to study whether the missingness is ignorable, and if not, what is the direction of the bias. The characteristics of AR people who cannot be linked to ACS people and thus may be missed by the ACS can shed light on the demographic biases of ACS omissions, and vice versa.

[78] In a similar spirit, we display results from regressions estimated on ACS responses to adjust the AR in nonresponding units in Table 3.13E.

[79] The ACS sample and base weights may also contribute to the undercount. Brown et al. (2023) show that administrative records produce a total population estimate 6.5 million under the 2020 Census when restricted to the 2020 Census housing structure collection universe, but the administrative record estimate is 7.8 million higher than the 2020 Census when using all administrative record addresses in the United States.

Table 3.8A ACS and AR Percent of People by Age and Race/Ethnicity, Responding ACS Housing Units, Separately for Linked vs. Unlinked Records

| | ACS, Linked People | ACS, Unlinked, Has PIK, | ACS, Unlinked, No PIK | AR, Linked People | AR, Unlinked People in ACS HUs with Person Responses | AR, Unlinked People in Vacant or Delete ACS HUs |
|---|---|---|---|---|---|---|
| Percent of Source (ACS or AR) Total | 86.92 | 8.08 | 5.00 | 69.34 | 27.64 | 3.02 |
| | Percent by Age | | | | | |
| Age 0-17 | 20.73 | 21.06 | 27.70 | 21.28 | 26.40 | 27.98 |
| Age 18+ | 79.27 | 78.92 | 72.30 | 78.73 | 73.60 | 72.02 |
| | Percent of Age 0-17 | | | | | |
| Non-Hispanic | 81.55 | 77.06 | 67.95 | 81.36 | 72.16 | 74.95 |
| NH AIAN Alone | 0.59 | 2.44 | 1.76 | 0.78 | 1.57 | 2.99 |
| NH Asian Alone | 5.64 | 4.65 | 9.19 | 5.66 | 3.63 | 2.43 |
| NH Black Alone | 8.70 | 10.48 | 13.89 | 7.68 | 15.05 | 20.97 |
| NH NHPI Alone | 0.13 | 0.35 | 0.36 | 0.13 | 0.25 | 0.24 |
| NH White Alone | 61.12 | 53.58 | 39.57 | 62.33 | 48.27 | 45.02 |
| NH AIAN & White | 0.70 | 0.93 | 0.51 | 0.38 | 0.18 | 0.17 |
| NH Asian & White | 1.98 | 1.49 | 0.83 | 1.90 | 0.92 | 0.72 |
| NH Black & White | 1.95 | 1.89 | 1.18 | 1.72 | 1.61 | 1.76 |
| NH AIAN and Black | 0.07 | 0.11 | 0.09 | 0.05 | 0.04 | 0.06 |
| NH Other Multi-Race | 0.66 | 1.14 | 0.56 | 0.73 | 0.64 | 0.60 |
| Hispanic | 18.47 | 22.94 | 32.05 | 18.65 | 27.85 | 25.02 |
| | Percent of Age 18+ | | | | | |
| Non-Hispanic | 89.06 | 85.24 | 68.98 | 88.92 | 79.12 | 82.91 |
| NH AIAN Alone | 0.46 | 1.62 | 1.28 | 0.71 | 1.16 | 2.22 |
| NH Asian Alone | 5.82 | 6.74 | 9.27 | 5.57 | 6.54 | 4.22 |
| NH Black Alone | 8.04 | 9.74 | 11.20 | 7.60 | 12.14 | 16.76 |
| NH NHPI Alone | 0.11 | 0.29 | 0.32 | 0.13 | 0.20 | 0.20 |
| NH White Alone | 72.97 | 64.57 | 45.46 | 73.58 | 57.39 | 58.04 |
| NH AIAN & White | 0.54 | 0.70 | 0.38 | 0.28 | 0.15 | 0.19 |
| NH Asian & White | 0.48 | 0.57 | 0.38 | 0.43 | 0.55 | 0.36 |
| NH Black & White | 0.32 | 0.37 | 0.28 | 0.26 | 0.44 | 0.42 |
| NH AIAN and Black | 0.06 | 0.08 | 0.08 | 0.04 | 0.04 | 0.05 |
| NH Other Multi-Race | 0.26 | 0.55 | 0.33 | 0.32 | 0.49 | 0.45 |
| Hispanic | 10.94 | 14.76 | 31.02 | 11.08 | 20.88 | 17.09 |

Notes: Here we use the pre-disclosure avoidance 2016-2020 ACS responding housing units with base weights. The linked people are linked by PIK in the same ACS housing unit. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.8B Percent Citizens by Age and Race/Ethnicity, Responding ACS Housing Units, Separately for Linked vs. Unlinked Records

| | ACS, Linked People | ACS, Unlinked, Has PIK, | ACS, Unlinked, No PIK | AR, Linked People | AR, Unlinked People in ACS HUs with Person Responses | AR, Unlinked People in Vacant or Delete ACS HUs |
|---|---|---|---|---|---|---|
| Total | 95.85 | 92.70 | 74.62 | 95.91 | 88.02 | 90.92 |
| | Age Categories | | | | | |
| Age 0-17 | 98.79 | 96.86 | 84.55 | 98.75 | 98.06 | 98.07 |
| Age 18+ | 95.09 | 91.61 | 70.82 | 95.13 | 84.41 | 88.15 |
| | Age 0-17 | | | | | |
| Non-Hispanic | 98.96 | 97.84 | 86.37 | 98.91 | 98.29 | 98.34 |
| NH AIAN Alone | 99.90 | 99.70 | 99.60 | 99.41 | 99.18 | 99.46 |
| NH Asian Alone | 93.28 | 86.34 | 42.42 | 93.84 | 90.39 | 90.91 |
| NH Black Alone | 98.01 | 96.79 | 94.81 | 97.91 | 98.33 | 98.70 |
| NH NHPI Alone | 92.90 | 93.72 | 91.52 | 93.99 | 93.67 | 92.16 |
| NH White Alone | 99.60 | 98.92 | 92.56 | 99.49 | 98.91 | 98.61 |
| NH AIAN & White | 100.0 | 99.76 | 99.54 | 99.72 | 98.87 | 98.67 |
| NH Asian & White | 98.56 | 97.21 | 76.39 | 98.48 | 95.46 | 95.01 |
| NH Black & White | 99.89 | 99.27 | 97.17 | 99.71 | 98.88 | 98.69 |
| NH AIAN and Black | 99.96 | 99.11 | 99.03 | 99.38 | 98.24 | 98.90 |
| NH Other Multi-Race | 99.45 | 98.79 | 94.83 | 98.78 | 96.52 | 96.19 |
| Hispanic | 97.95 | 93.62 | 80.69 | 98.00 | 97.44 | 97.42 |
| | Age 18+ | | | | | |
| Non-Hispanic | 96.80 | 94.86 | 85.50 | 96.90 | 91.52 | 92.63 |
| NH AIAN Alone | 99.48 | 99.63 | 98.95 | 97.77 | 95.89 | 97.38 |
| NH Asian Alone | 73.94 | 66.52 | 44.61 | 76.42 | 59.57 | 60.30 |
| NH Black Alone | 96.29 | 94.11 | 89.71 | 96.54 | 93.67 | 95.38 |
| NH NHPI Alone | 84.48 | 85.19 | 80.26 | 87.47 | 76.96 | 73.60 |
| NH White Alone | 98.68 | 97.79 | 92.46 | 98.54 | 94.78 | 94.18 |
| NH AIAN & White | 99.82 | 99.65 | 98.65 | 98.71 | 95.82 | 97.05 |
| NH Asian & White | 93.86 | 91.89 | 62.22 | 93.34 | 87.59 | 84.23 |
| NH Black & White | 98.10 | 97.01 | 92.35 | 97.49 | 96.34 | 96.31 |
| NH AIAN and Black | 99.55 | 98.89 | 98.56 | 98.34 | 94.88 | 96.25 |
| NH Other Multi-Race | 96.11 | 96.50 | 88.02 | 91.09 | 80.93 | 82.83 |
| Hispanic | 81.17 | 72.85 | 38.19 | 81.00 | 57.50 | 66.36 |

Notes: Here we use the pre-disclosure avoidance 2016-2020 ACS responding housing units with base weights. The linked people are linked by PIK in the same ACS housing unit. The ACS citizenship values are as-reported ACS citizenship when available. If as-reported ACS citizenship is not available, and AR citizenship is available, then AR citizenship is used. Otherwise, modeled predictions are used. AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

In Tables 3.8A-B we calculate ACS and AR estimates separately for linked and unlinked people in the responding units. Comparing the results for linked people shows how much discrepancies in assignment of age and race/ethnicity and the uncertainty over whether the AR people are residents in

the ACS housing units matter. Table 3.8A illustrates that among linked people, the single-race non-Hispanic AIAN and non-Hispanic White shares are higher in AR, while the two-race non-Hispanic AIAN and White share is higher in the ACS. This could reflect greater ACS reporting of the AIAN and White multi-race category in more recent years. The categories including non-Hispanic Black alone or in combination have higher shares in the ACS. Otherwise, the estimates among linked people are similar, suggesting that the age and race/ethnicity assignments are comparable and that the AR person-place probabilities do not affect the shares much.

AR could omit a tabulation month ACS housing unit resident either because the person's ARs lack that address, or because the person does not appear in the set of AR with PIKs. Unlinked ACS people with a PIK should be comparable to the AR group without their tabulation month residence in AR. Some of the people without PIKs in the ACS are not in the AR with PIK universe, so they should be a suitable comparison group to the people not appearing in AR with PIKs.[80] In the table we thus divide the unlinked ACS people into ones with vs. without a PIK. The ACS may omit a person either when a household roster response is incomplete, or in a housing unit classified as a vacant or delete in error. We therefore divide unlinked AR people into ones who are in housing units containing ACS people vs. those in housing units classified by the ACS as vacant or a delete.[81]

The unlinked shares of non-Hispanic Whites are much lower than the linked shares. The changes in the distribution from linked to unlinked people are quite similar in the ACS and AR. The AR distributions are not very different for those in housing units with ACS people vs. ACS vacant or delete housing units. What is driving the difference in ACS vs. AR demographic estimates in the responding ACS housing unit universe is that the linked share of people is so much higher for the ACS (86.92 percent) than for AR (69.34 percent) in a context where the age and race/ethnicity distributions are dramatically different for linked vs. unlinked.

Citizen share estimates for linked and unlinked people are shown in Table 3.8B. Unlinked people have lower estimated citizen shares, especially among ACS people without a PIK. The ACS and AR missingness in responding ACS housing units thus appears to be nonignorable with respect to citizenship, upwardly biasing the estimates. The citizen share gap is most dramatic for voting-age Hispanics: 81.17 percent are citizens among linked ACS people vs. 38.16 percent among those without a PIK. The difference among non-Hispanic Asians is also large.

The comparability of the changes in age-race/ethnicity and citizen shares across linked and unlinked groups supports the possibility that some of the ACS and AR unlinked people are the same people. The people in AR who may be missing from the ACS appear to be disproportionately Hispanic noncitizens, which is consistent with the results showing a higher AR voting-age Hispanic share (Table 3.6) and a lower voting-age Hispanic citizen share (Table 3.7A). Significant shares of the people in the ACS unlinked to AR for either reason (no PIK or the ACS address is not in AR) are voting-age Hispanic noncitizens. Our preferred person-place probability normalization helps mitigate the loss of ACS housing unit residents whose ACS address is missing from their AR by placing a higher weight (compared to the other methods for using person-place probabilities) on people who are more likely to

[80] Some ACS people without a PIK may be in AR. The absence of a PIK could be due to discrepancies between their PII reported in the ACS vs. AR, or their PII is not reported in the ACS.

[81] The ACS has a two-month minimum residency eligibility rule. Note that AR could potentially include people who lived in the ACS housing unit for less than two months.

be missing their tabulation month residence in AR. But none of the AR-based methods account for people absent from the AR PIK universe, a group that the ACS suggests has an even lower voting-age Hispanic citizen share than the group missing their ACS address in AR. *This raises the prospect that the AR-based methods used here overestimate the voting-age Hispanic citizen share, despite producing lower estimates than the methods relying more on ACS responses.*[82]

Table 3.9 Citizen Shares by Source in Responding ACS Housing Units

| | % Citizens, ACS | % Citizens, AR and Modeling in Place of ACS Edits and Imputes | % of ACS Sample | % Citizens, AR | % of AR Sample |
|---|---|---|---|---|---|
| | Voting-Age Hispanic | | | | |
| As Reported, AR | 82.09 | 82.09 | 69.14 | 81.27 | 47.35 |
| As Reported, Modeled | 56.46 | 56.46 | 0.05 | 79.53 | 0.04 |
| Edited, AR | 86.21 | 77.65 | 0.50 | 77.80 | 0.35 |
| Edited, Modeled | 89.63 | 81.25 | 0.00 | 82.26 | 0.00 |
| Imputed, AR | 75.94 | 68.87 | 6.86 | 68.41 | 4.69 |
| Imputed, Modeled | 72.49 | 77.84 | 0.01 | 78.23 | 0.01 |
| As Reported, No Link, Has PIK | 76.07 | 76.07 | 10.39 | | |
| Edited, No Link, Has PIK | 83.90 | 81.92 | 0.07 | | |
| Imputed, No Link, Has PIK | 74.87 | 79.30 | 1.28 | | |
| As Reported, No Link, No PIK | 36.95 | 36.95 | 9.49 | | |
| Edited, No Link, No PIK | 65.62 | 52.48 | 0.05 | | |
| Imputed, No Link, No PIK | 62.52 | 43.32 | 2.14 | | |
| AR, Occupied ACS HU | | | | 60.70 | 43.78 |
| Modeled, Occupied ACS HU | | | | 83.98 | 0.32 |
| AR, Vacant/Delete ACS HU | | | | 66.24 | 3.46 |
| Modeled, Vacant/Delete ACS HU | | | | 85.81 | 0.02 |
| Total | 76.24 | 75.33 | 100.0 | 71.15 | 100.0 |
| | Voting-Age Non-Hispanic | | | | |
| As Reported, AR | 96.91 | 96.91 | 83.20 | 97.08 | 69.06 |
| As Reported, Modeled | 87.09 | 87.09 | 0.07 | 96.63 | 0.06 |
| Edited, AR | 94.84 | 97.46 | 0.26 | 97.60 | 0.22 |
| Edited, Modeled | 99.06 | 96.18 | 0.00 | 96.24 | 0.00 |
| Imputed, AR | 95.59 | 94.87 | 4.17 | 94.99 | 3.34 |
| Imputed, Modeled | 95.34 | 96.11 | 0.01 | 96.19 | 0.01 |

[82] Brown et al (2023) estimate a lower 2020 voting-age Hispanic citizen share (69.16 when including only people with SSNs or ITINs, and 66.02 percent when also including people without SSNs or ITINs) compared to the 2020 AR estimate of 71.23 percent in Table 3.5F. They include administrative records for any U.S address.

| | | | | | |
|---|---|---|---|---|---|
| As Reported, No Link, Has PIK | 94.92 | 94.92 | 8.03 | | |
| Edited, No Link, Has PIK | 91.59 | 94.77 | 0.03 | | |
| Imputed, No Link, Has PIK | 94.60 | 94.28 | 0.62 | | |
| As Reported, No Link, No PIK | 85.64 | 85.64 | 2.94 | | |
| Edited, No Link, No PIK | 86.97 | 84.39 | 0.01 | | |
| Imputed, No Link, No PIK | 92.68 | 84.88 | 0.64 | | |
| AR, Occupied ACS HU | | | | 91.40 | 24.51 |
| Modeled, Occupied ACS HU | | | | 91.15 | 0.08 |
| AR, Vacant/Delete ACS HU | | | | 92.65 | 2.70 |
| Modeled, Vacant/Delete ACS HU | | | | 91.53 | 0.01 |
| Total | 96.30 | 96.22 | 100.0 | 95.50 | 100.0 |

Notes: The samples are people in housing units that responded in the 2016-2020 ACS. AR in columns 4 and 5 uses household rosters from administrative record sources, and their citizenship is from numerous survey and administrative record sources and modeling. AR in the row headings means AR citizenship is available for the person. Modeled in the row headings means AR citizenship is not available for the person, but a modeled value is available. Edited (imputed) means the ACS citizenship value is edited (imputed). No link means the ACS person record cannot be linked to an AR record in the same housing unit by PIK. Occupied ACS HU means the ACS response includes people. Vacant/delete ACS HU means the ACS classified the housing unit as a vacant or delete. The pre-disclosure avoidance ACS file is used here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

We conduct detailed analysis of the different components of the Hispanic and voting-age non-Hispanic citizen share estimate for responding ACS housing units in Table 3.9. The first column for the ACS uses ACS household rosters and ACS citizenship. The second column uses ACS rosters, ACS as reported citizenship, and AR/modeled citizenship in place of ACS edits and imputations. The fourth column uses AR household rosters and AR/modeled citizenship. Sample shares are provided to show the relative importance of each component.

Replacing ACS edits and imputations with AR/modeling reduces the voting-age Hispanic citizen share estimate by 0.91 percentage points. The biggest contributors to the difference are replacing ACS imputations with AR citizenship (7.07 percentage points lower citizen share for 6.86 percent of the ACS sample) and replacing imputations with modeled citizenship for the ACS people lacking PIKs (19.2 percentage points lower for 2.14 percent of the ACS sample). AR in place of as-reported ACS citizenship lowers the citizen share somewhat (0.82 percentage points lower for 47.35 percent of the AR sample), as does placing AR-sourced people in housing units classified as vacant or a delete in the ACS – the 66.24 percent citizen estimate for the group with AR citizenship in these housing units is well below the average. The largest difference comes from occupied ACS housing units where AR-sourced people cannot be linked to people in the ACS household rosters by PIK. Though the unlinked ACS people have a lower citizen share than the unlinked AR people in those housing units, the AR person share in this group is more than twice as large, and its citizen share is below average. Some of the unlinked people

are surely the same people in the ACS and AR, but there is a large excess group of unlinked AR people with a relatively low estimated citizen share.

There are sizable groups of unlinked ACS and AR people in the same housing units for voting-age non-Hispanics as well, but the relative size of the unlinked groups is smaller, and citizen share estimates for them are less dramatically different from those of the linked groups, so the citizen share estimate is little affected by the unlinked groups.

Table 3.10 Percent ITINs Among AR Voting-Age Noncitizens

| | Linked to ACS Person | Linked to Occupied ACS Housing Unit, But No Link to ACS Person | Vacant/Delete ACS Housing Unit | Nonresponding ACS Housing Unit |
|---|---|---|---|---|
| NH Asian Alone | 1.97 | 12.29 | 5.92 | 6.44 |
| NH Black Alone | 0.74 | 3.11 | 2.91 | 2.52 |
| NH White Alone | 1.60 | 10.75 | 9.61 | 11.18 |
| Hispanic | 15.88 | 55.72 | 44.34 | 46.08 |

Notes: The samples are AR noncitizens age 18 and over in responding 2016-2020 ACS housing units. An ITIN (Individual Taxpayer Identification Number) is a nine-digit number in the SSN field that is in a certain publicly disclosed numeric range. The IRS issues ITINs to people needing to correspond with the IRS who are ineligible to have an SSN. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

ACS responses could omit voting-age Hispanic noncitizens at higher rates than other groups. Respondents may fail to list noncitizens, particularly ones with unknown legal status. We can distinguish noncitizens with SSNs from noncitizens with ITINs in the administrative data. People with SSNs had authorization to work at some point in their life, while those with ITINs never have had authorization.[83] Table 3.10 shows the percent of AR voting-age noncitizens with ITINs, separately for those linked to a person in the ACS roster for the same housing unit, those in an occupied ACS housing unit with no link to anyone in the ACS roster, people in a housing unit that is classified as vacant or a delete by the ACS, and those in a nonresponding ACS housing unit. For all the major race/ethnicity groups, ITINs make up the highest share of noncitizens in occupied ACS housing units where the person cannot be linked to an ACS person, and their share of noncitizens is lowest for people who are linked to an ACS person. The patterns across race/ethnicity groups are similar, and the levels are far higher among Hispanics than the other groups. These patterns are consistent with the possibility that people with unknown immigration status are more likely to be left off the ACS household roster. A higher share of Hispanics has unknown immigration status, hence this effect is most evident in that group.

[83] Nonetheless, some people with ITINs work and by law owe taxes. The National Immigration Law Center (2017) explains that immigrants with ITINs file tax returns not only because they are required to do so, but to prove "good moral "character" and document U.S. work history and physical presence if given an opportunity to gain legal status in the future. Undocumented immigrants with ITINs should thus be more likely than other undocumented immigrants to be long-term U.S. residents.

Table 3.11 Mean Number of AR Addresses Per Person Among Voting-Age Population by Race/Ethnicity

| | Citizens | Noncitizens | Noncitizens with SSNs | Noncitizens with ITINs | All |
|---|---|---|---|---|---|
| NH Asian Alone | 2.65 | 2.42 | 2.47 | 1.43 | 2.58 |
| NH Black Alone | 3.57 | 2.86 | 2.88 | 1.60 | 3.54 |
| NH White Alone | 2.86 | 2.40 | 2.46 | 1.47 | 2.85 |
| Hispanic | 3.15 | 2.27 | 2.47 | 1.90 | 2.92 |

Notes: The samples are people age 18 and over in particular race/ethnicity groups that include housing units in the initial 2016-2020 ACS among their AR MAFIDs (addresses). The observations are weighted by ACS housing unit base weights. Normalized person-place probabilities are not included in the weights here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Another potential explanation is that administrative records contain greater error in placing voting-age Hispanic noncitizens in the correct residences. They may be more mobile than other groups, leading to more AR addresses and greater uncertainty about which one (if any) is their ACS tabulation month residence. Alternatively, they could have fewer AR addresses due to less use of public programs. In that case, AR may not capture all their residences, which could also cause placement error. Voting-age Hispanic citizens have the second highest number of AR addresses among citizen-race/ethnicity groups and the lowest among noncitizens (Table 3.11). People with ITINs have by far the fewest number of AR addresses. Hispanics with ITINs have more than the other race/ethnicity groups, however.

The smaller number of AR addresses among noncitizens could indicate less complete coverage of noncitizens' migration compared to that of citizens, leading to more placement errors. And as discussed in Section 2.7, the person-place weight normalization may exacerbate placement errors for part-time U.S. residents, many of whom are likely to be noncitizens. The normalization could thus contribute to an AR overestimate of noncitizens.[84]

Table 3.12 Comparison of Person-Place Probabilities, Normalized Person-Place Probabilities, and Actual Person-Place Linkage Rates

| | Person-Place Linkage Rate | Mean Person-Place Probability | Mean Normalized Person-Place Probability |
|---|---|---|---|
| | Voting-Age Hispanic | | |
| Citizens | 40.33 | 45.98 | 50.10 |
| Noncitizens | 30.27 | 45.38 | 67.74 |
| With SSNs | 38.41 | 45.60 | 60.06 |

[84] Considering the heightened immigration enforcement at the U.S.-Mexico border during this period, it is unlikely that many Hispanic undocumented immigrants would risk going out and back into the United States, though. Circular migration is more common when border enforcement is less strict (see Massey et al. (2014)). Using the ACS, Warren (2021) estimates that the Mexican undocumented population in the United States fell from 6.6 to 4.8 million people between 2010 and 2019. The estimated decline could reflect increased reluctance by Hispanic undocumented immigrants to respond to the ACS, given more active interior immigration enforcement, however.

| | | | |
|---|---|---|---|
| With ITINs | 14.66 | 44.94 | 82.53 |
| | Voting-Age NH Asian | | |
| Citizens | 50.99 | 54.58 | 58.16 |
| Noncitizens | 43.87 | 47.20 | 63.70 |
| With SSNs | 45.23 | 47.43 | 62.71 |
| With ITINs | 17.06 | 42.59 | 83.43 |
| | Voting-Age NH Black | | |
| Citizens | 34.65 | 42.72 | 45.62 |
| Noncitizens | 33.09 | 42.20 | 55.40 |
| With SSNs | 33.37 | 42.28 | 55.05 |
| With ITINs | 15.45 | 37.54 | 77.81 |
| | Voting-Age NH White | | |
| Citizens | 51.18 | 52.53 | 56.23 |
| Noncitizens | 35.17 | 42.93 | 63.58 |
| With SSNs | 36.65 | 43.05 | 62.47 |
| With ITINs | 10.40 | 40.93 | 81.91 |

Notes: An ITIN (Individual Taxpayer Identification Number) is a nine-digit number in the SSN field that is in a certain publicly disclosed numeric range. The IRS issues ITINs to people needing to correspond with the IRS who are ineligible to have an SSN. The person-place probability is the probability that the AR person is located at the ACS housing unit in the tabulation month, without normalization. AR people are classified as being linked only if linked to people in the ACS housing units where they have a positive probability of being located. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.12 compares the actual person-place linkage rates among people in AR and ACS household rosters for the same housing units to the mean person-place probabilities from the person-place model and the normalized mean person-place probabilities. The mean person-place probabilities are not very different across the different citizenship status groups, varying by just a percentage point for Hispanics, suggesting similar AR address quality. The mean predicted probabilities are always higher than the actual person-place linkage rates, and normalized rates are even higher. Among citizens, the three rates vary about half as much for non-Hispanic Whites as they do for Hispanics and non-Hispanic Blacks, which could reflect both higher AR coverage of non-Hispanic White addresses and a lower propensity to omit them in ACS rosters. The gaps between the actual person-place linkage rates and the mean person-place probabilities are not much bigger for noncitizens with SSNs than they are for citizens, suggesting that the omission rates for noncitizens with SSNs from the ACS rosters are not significantly greater than for citizens. Omission rates appear to be much higher for people with ITINs, as reflected by the fact that the differences between the person-place linkage rates and the mean person-place probabilities are three or more times as large as they are for the other two citizenship groups. The lower number of AR addresses among noncitizens displayed in Table 3.11 shows up here as a large gap between the initial and normalized person-place probabilities, especially among people with ITINs. The total gaps between actual linkage rates and normalized probabilities for people with ITINs are enormous, which should lead to significant uncertainty for their estimates. This occurs in all the race/ethnicity groups. As shown in Table 3.10, ITINS are only an important share of noncitizens for Hispanics, so their estimates are most affected by this issue. The greater propensity for ITINs to be included in AR, but not the ACS is consistent with Brown et al. (2023), who show that noncitizens with unknown legal status make up a significant share of those included in 2020 AR population estimates, but not the 2020 Census or 2020 PEP estimates.

Table 3.13A AR Population by Age and Race/Ethnicity in Responding 2020 ACS Housing Units, No Person-Place Probability Normalization

| | Responding HUs | Nonresponding HUs |
|---|---|---|
| Total Population | 120,800,000 | 155,100,000 |
| | Percent by Age | |
| Age 0-17 | 18.29 | 22.93 |
| Age 18+ | 81.71 | 77.07 |
| | Percent of Age 0-17 | |
| Non-Hispanic | 79.00 | 71.00 |
| NH AIAN Alone | 1.06 | 0.89 |
| NH Asian Alone | 4.47 | 3.25 |
| NH Black Alone | 10.64 | 17.14 |
| NH NHPI Alone | 0.18 | 0.22 |
| NH White Alone | 59.20 | 46.27 |
| NH AIAN & White | 0.16 | 0.14 |
| NH Asian & White | 1.20 | 0.82 |
| NH Black & White | 1.47 | 1.65 |
| NH AIAN and Black | 0.03 | 0.04 |
| NH Other Multi-Race | 0.59 | 0.58 |
| Hispanic | 20.99 | 29.00 |
| | Percent of Age 18+ | |
| Non-Hispanic | 87.44 | 80.59 |
| NH AIAN Alone | 0.79 | 0.77 |
| NH Asian Alone | 5.51 | 4.95 |
| NH Black Alone | 8.90 | 15.04 |
| NH NHPI Alone | 0.15 | 0.19 |
| NH White Alone | 70.77 | 58.16 |
| NH AIAN & White | 0.14 | 0.14 |
| NH Asian & White | 0.48 | 0.41 |
| NH Black & White | 0.33 | 0.46 |
| NH AIAN and Black | 0.03 | 0.05 |
| NH Other Multi-Race | 0.36 | 0.42 |
| Hispanic | 12.57 | 19.41 |

Notes: AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The person-place probabilities used in the person weights are not normalized here. These estimates are otherwise generated in the same way as those in the last column of Tables 3.4A and 3.4E. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13B AR Estimates for Percent of Population in Nonresponding 2020 ACS Housing Units and Citizen Shares by Age and Race/Ethnicity, No Person-Place Probability Normalization

| | Percent of Population in Nonresponding 2020 ACS HUs | Percent Citizens in Responding ACS HUs | Percent Citizens in Nonresponding ACS HUs | Percent Citizens in All ACS HUs |
|---|---|---|---|---|
| Total | 56.22 | 94.99 | 92.78 | 93.75 |
| | Age Categories | | | |
| Age 0-17 | 61.67 | 98.66 | 98.42 | 98.53 |
| Age 18+ | 54.77 | 94.17 | 91.09 | 92.48 |
| | Age 0-17 | | | |
| Non-Hispanic | 59.11 | 98.80 | 98.53 | 98.64 |
| NH AIAN Alone | 57.34 | 99.08 | 98.78 | 98.91 |
| NH Asian Alone | 53.90 | 95.55 | 94.33 | 94.89 |
| NH Black Alone | 72.17 | 98.62 | 98.85 | 98.82 |
| NH NHPI Alone | 67.22 | 95.81 | 94.46 | 94.88 |
| NH White Alone | 55.70 | 99.12 | 98.75 | 98.92 |
| NH AIAN & White | 59.67 | 98.93 | 98.62 | 98.77 |
| NH Asian & White | 52.49 | 97.37 | 96.29 | 96.77 |
| NH Black & White | 64.27 | 99.03 | 98.98 | 98.96 |
| NH AIAN and Black | 70.22 | 97.74 | 98.85 | 98.50 |
| NH Other Multi-Race | 61.16 | 97.25 | 96.99 | 97.09 |
| Hispanic | 68.99 | 98.22 | 98.16 | 98.19 |
| | Age 18+ | | | |
| Non-Hispanic | 52.74 | 96.23 | 95.02 | 95.59 |
| NH AIAN Alone | 54.23 | 96.91 | 95.54 | 96.20 |
| NH Asian Alone | 52.09 | 78.71 | 75.10 | 76.82 |
| NH Black Alone | 67.18 | 96.47 | 96.22 | 96.32 |
| NH NHPI Alone | 60.66 | 83.91 | 80.90 | 82.07 |
| NH White Alone | 49.87 | 97.64 | 96.47 | 97.06 |
| NH AIAN & White | 55.37 | 97.90 | 97.29 | 97.56 |
| NH Asian & White | 50.79 | 93.37 | 90.88 | 92.10 |
| NH Black & White | 62.63 | 97.82 | 97.65 | 97.68 |
| NH AIAN and Black | 66.18 | 96.57 | 96.99 | 96.89 |
| NH Other Multi-Race | 59.03 | 89.57 | 88.04 | 88.69 |
| Hispanic | 65.14 | 79.77 | 74.77 | 76.52 |

Notes: AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The person-place probabilities used in the person weights are not normalized here. These estimates are otherwise generated in the same way as those in the last column of Tables 3.4F, 3.5A, 3.5E, and 3.5F. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13C AR Population by Age and Race/Ethnicity, Responding 2020 ACS Housing Units, Alternative Person-Place Probability Normalization

| | Responding HUs | Nonresponding HUs |
|---|---|---|
| Total Population | 114,200,000 | 146,600,000 |
| | Percent by Age | |
| Age 0-17 | 20.80 | 25.27 |
| Age 18+ | 79.20 | 74.73 |
| | Percent of Age 0-17 | |
| Non-Hispanic | 79.98 | 71.62 |
| NH AIAN Alone | 0.85 | 0.74 |
| NH Asian Alone | 4.71 | 3.49 |
| NH Black Alone | 9.45 | 15.44 |
| NH NHPI Alone | 0.15 | 0.19 |
| NH White Alone | 61.39 | 48.53 |
| NH AIAN & White | 0.15 | 0.14 |
| NH Asian & White | 1.27 | 0.89 |
| NH Black & White | 1.43 | 1.60 |
| NH AIAN and Black | 0.03 | 0.04 |
| NH Other Multi-Race | 0.57 | 0.57 |
| Hispanic | 20.02 | 28.38 |
| | Percent of Age 18+ | |
| Non-Hispanic | 87.15 | 79.83 |
| NH AIAN Alone | 0.62 | 0.61 |
| NH Asian Alone | 5.71 | 5.13 |
| NH Black Alone | 8.41 | 14.03 |
| NH NHPI Alone | 0.13 | 0.16 |
| NH White Alone | 71.02 | 58.51 |
| NH AIAN & White | 0.13 | 0.14 |
| NH Asian & White | 0.47 | 0.40 |
| NH Black & White | 0.31 | 0.43 |
| NH AIAN and Black | 0.03 | 0.04 |
| NH Other Multi-Race | 0.33 | 0.39 |
| Hispanic | 12.85 | 20.17 |

Notes: AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The person-place probability normalization here adds the probability of not having the person's ACS tabulation month residence in their AR to the denominator. These estimates are otherwise generated in the same way as those in the last column of Tables 3.4A and 3.4E. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13D AR Estimates for Percent of Population in Nonresponding 2020 ACS Housing Units and Citizen Shares by Age and Race/Ethnicity, Alternative Person-Place Probability Normalization

| | Percent of Population in Nonresponding 2020 ACS HUs | Percent Citizens in Responding ACS HUs | Percent Citizens in Nonresponding ACS HUs | Percent Citizens in All ACS HUs |
|---|---|---|---|---|
| Total | 56.21 | 94.22 | 91.61 | 92.79 |
| | Age Categories | | | |
| Age 0-17 | 60.95 | 98.18 | 99.00 | 99.01 |
| Age 18+ | 54.79 | 92.98 | 89.05 | 90.85 |
| | Age 0-17 | | | |
| Non-Hispanic | 58.29 | 99.24 | 99.06 | 99.06 |
| NH AIAN Alone | 57.49 | 99.30 | 99.06 | 99.17 |
| NH Asian Alone | 53.61 | 97.32 | 96.52 | 96.91 |
| NH Black Alone | 71.83 | 99.00 | 99.21 | 99.12 |
| NH NHPI Alone | 66.45 | 96.74 | 95.53 | 95.94 |
| NH White Alone | 55.23 | 99.42 | 99.22 | 99.23 |
| NH AIAN & White | 59.33 | 99.33 | 99.14 | 99.20 |
| NH Asian & White | 52.22 | 98.48 | 97.80 | 98.18 |
| NH Black & White | 63.66 | 99.40 | 99.24 | 99.30 |
| NH AIAN and Black | 69.60 | 98.66 | 99.07 | 98.95 |
| NH Other Multi-Race | 60.79 | 98.19 | 98.05 | 98.10 |
| Hispanic | 68.87 | 98.88 | 98.86 | 98.89 |
| | Age 18+ | | | |
| Non-Hispanic | 52.60 | 95.56 | 94.05 | 94.80 |
| NH AIAN Alone | 54.45 | 95.45 | 93.44 | 94.40 |
| NH Asian Alone | 52.11 | 76.22 | 72.28 | 74.15 |
| NH Black Alone | 66.89 | 95.86 | 95.48 | 95.60 |
| NH NHPI Alone | 60.55 | 79.04 | 76.06 | 77.24 |
| NH White Alone | 49.96 | 97.20 | 95.71 | 96.49 |
| NH AIAN & White | 55.36 | 97.45 | 96.80 | 97.09 |
| NH Asian & White | 50.77 | 92.33 | 89.37 | 90.81 |
| NH Black & White | 62.25 | 97.31 | 97.10 | 97.15 |
| NH AIAN and Black | 65.75 | 95.95 | 96.24 | 96.13 |
| NH Other Multi-Race | 59.11 | 86.71 | 84.40 | 85.34 |
| Hispanic | 65.54 | 75.20 | 69.25 | 71.32 |

Notes: AR here means that the household rosters come from administrative record sources, and their demographics are from numerous survey and administrative record sources and modeling. The person-place probability normalization here adds the probability of not having the person's ACS tabulation month residence in their AR to the denominator. These estimates are otherwise generated in the same way as those in the last column of Tables 3.4F, 3.5A, 3.5E, and 3.5F. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13E AR Population by Age and Race/Ethnicity in Nonresponding 2020 ACS Housing Units, OLS Models Trained on ACS

| | Normalized Without Probability Residence Not in AR | Normalized With Probability Residence Not in AR |
|---|---|---|
| Total Population | 172,500,000 | 172,900,000 |
| | Percent by Age | |
| Age 0-17 | 23.72 | 23.81 |
| Age 18+ | 76.28 | 76.19 |
| | Percent of Age 0-17 | |
| Non-Hispanic | 74.17 | 74.15 |
| NH AIAN Alone | 1.14 | 1.18 |
| NH Asian Alone | 4.35 | 4.34 |
| NH Black Alone | 13.15 | 13.24 |
| NH NHPI Alone | 0.22 | 0.23 |
| NH White Alone | 51.14 | 51.01 |
| NH AIAN & White | 0.55 | 0.55 |
| NH Asian & White | 1.13 | 1.12 |
| NH Black & White | 1.77 | 1.76 |
| NH AIAN and Black | 0.08 | 0.08 |
| NH Other Multi-Race | 0.63 | 0.64 |
| Hispanic | 25.83 | 25.85 |
| | Percent of Age 18+ | |
| Non-Hispanic | 82.74 | 82.72 |
| NH AIAN Alone | 0.87 | 0.88 |
| NH Asian Alone | 5.50 | 5.49 |
| NH Black Alone | 12.35 | 12.44 |
| NH NHPI Alone | 0.18 | 0.18 |
| NH White Alone | 62.14 | 62.03 |
| NH AIAN & White | 0.46 | 0.46 |
| NH Asian & White | 0.43 | 0.43 |
| NH Black & White | 0.40 | 0.41 |
| NH AIAN and Black | 0.07 | 0.07 |
| NH Other Multi-Race | 0.34 | 0.34 |
| Hispanic | 17.26 | 17.28 |

Notes: The Ordinary Least Squares (OLS) models predict the count for each of the 44 citizenship-age-race/ethnicity groups separately. The first column's models use AR regressors using the person-place probability normalization that does not include the probability of not having the person's ACS tabulation month residence in their AR in the denominator. The second column uses AR regressors including this probability in the denominator. These estimates are otherwise generated in the same way as those in the last column of Table 3.4E. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13F AR Estimated Citizen Shares in Nonresponding 2020 ACS Housing Units, OLS Models Trained on ACS

| | Normalized Without Probability Residence Not in AR | Normalized With Probability Residence Not in AR |
|---|---|---|
| Total | 93.19 | 93.09 |
| | Age Categories | |
| Age 0-17 | 97.62 | 97.53 |
| Age 18+ | 91.83 | 91.69 |
| | Age 0-17 | |
| Non-Hispanic | 98.02 | 97.95 |
| NH AIAN Alone | 99.78 | 99.77 |
| NH Asian Alone | 87.72 | 87.29 |
| NH Black Alone | 98.33 | 98.26 |
| NH NHPI Alone | 93.36 | 93.34 |
| NH White Alone | 98.76 | 98.69 |
| NH AIAN & White | 99.96 | 99.98 |
| NH Asian & White | 96.40 | 96.21 |
| NH Black & White | 99.59 | 99.52 |
| NH AIAN and Black | 99.76 | 99.77 |
| NH Other Multi-Race | 98.84 | 98.76 |
| Hispanic | 96.45 | 96.33 |
| | Age 18+ | |
| Non-Hispanic | 95.50 | 95.32 |
| NH AIAN Alone | 99.43 | 99.44 |
| NH Asian Alone | 70.77 | 70.63 |
| NH Black Alone | 95.91 | 95.85 |
| NH NHPI Alone | 83.15 | 83.16 |
| NH White Alone | 97.55 | 97.37 |
| NH AIAN & White | 99.67 | 99.75 |
| NH Asian & White | 90.54 | 90.29 |
| NH Black & White | 97.75 | 97.67 |
| NH AIAN and Black | 99.30 | 99.25 |
| NH Other Multi-Race | 95.22 | 95.16 |
| Hispanic | 74.26 | 74.28 |

Notes: The Ordinary Least Squares (OLS) models predict the count for each of the 44 citizenship-age-race/ethnicity groups separately. The first column's models use AR regressors using the person-place probability normalization that does not include the probability of not having the person's ACS tabulation month residence in their AR in the denominator. The second column uses AR regressors including this probability in the denominator. These estimates are otherwise generated the same as those in the last column of Table 3.5E. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.13G AR Population by Age and Race/Ethnicity in Nonresponding 2020 ACS Housing Units, OLS Models Trained on 2020 Census

| | Normalized Without Probability Residence Not in AR | Normalized With Probability Residence Not in AR |
|---|---|---|
| Total Population | 171,400,000 | 171,000,000 |
| | Percent by Age | |
| Age 0-17 | 23.64 | 23.68 |
| Age 18+ | 76.36 | 76.32 |
| | Percent of Age 0-17 | |
| Non-Hispanic | 72.85 | 72.90 |
| NH AIAN Alone | 1.08 | 1.10 |
| NH Asian Alone | 4.90 | 4.87 |
| NH Black Alone | 14.32 | 14.40 |
| NH NHPI Alone | 0.27 | 0.27 |
| NH White Alone | 46.58 | 46.57 |
| NH AIAN & White | 1.13 | 1.12 |
| NH Asian & White | 1.55 | 1.55 |
| NH Black & White | 2.06 | 2.07 |
| NH AIAN and Black | 0.13 | 0.13 |
| NH Other Multi-Race | 0.83 | 0.83 |
| Hispanic | 27.15 | 27.10 |
| | Percent of Age 18+ | |
| Non-Hispanic | 81.69 | 81.69 |
| NH AIAN Alone | 0.86 | 0.88 |
| NH Asian Alone | 5.83 | 5.82 |
| NH Black Alone | 12.98 | 13.03 |
| NH NHPI Alone | 0.21 | 0.21 |
| NH White Alone | 59.36 | 59.29 |
| NH AIAN & White | 0.99 | 0.99 |
| NH Asian & White | 0.49 | 0.49 |
| NH Black & White | 0.51 | 0.51 |
| NH AIAN and Black | 0.06 | 0.06 |
| NH Other Multi-Race | 0.44 | 0.44 |
| Hispanic | 18.31 | 18.31 |

Notes: The Ordinary Least Squares (OLS) models predict the count for each of the 22 age-race/ethnicity groups separately, using the 2020 Census housing unit population counts by age-race/ethnicity for responding ACS housing units as the dependent variables. The first column's models use AR regressors using the person-place probability normalization that does not include the probability of not having the person's ACS tabulation month residence in their AR in the denominator. The second column uses AR regressors including this probability in the denominator. These estimates are otherwise generated in the same way as those in the last column of Table 3.4E. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The results reported in Table 3.12 support the idea that the estimates are sensitive to how the person-place probabilities are used in the person weights. We study this by recalculating the AR results in Tables 3.4A, 3.4E, 3.4F 3.5A, 3.5E, and 3.5F using initial person-place probabilities without normalization (Tables 3.13A-B), as well as estimates using a normalization that adds to the denominator the probability that the person's tabulation month address is not one of their AR addresses (Tables 3.13C-D). The total population estimates in both responding and nonresponding housing units are lowest with the alternative normalization, and the estimates without normalization

are in between the other two. Both alternatives are much further away from the 2020 Census population estimates. The alternative normalization produces shares closer to the 2020 Census for some age-race/ethnicity categories, while the preferred normalization is closer for others. The preferred normalization has higher Hispanic shares. The preferred normalization's estimated citizen shares are lower than those from the other two methods. Consistent with the lower estimated citizen share for unlinked ACS PIKs vs. linked ones (columns 1 and 2 of Table 3.8B), these results illustrate that the preferred normalization's extra weight on people with a higher likelihood of not having their tabulation month residence in AR increases the influence of noncitizens in the estimates. This weighting is supported by the finding in columns 1 and 2 of Table 3.8B that absence of a person's tabulation month residence from their AR is negatively associated with being a citizen in the ACS.

Another way to estimate values for nonresponding units is by using models to predict what the ACS values would have been had there been responses. If AR household rosters are measured with more error than ACS ones, then using models to adjust the AR in nonresponding ACS housing units could reduce error. We estimated separate Ordinary Least Squares (OLS) regressions for each of the 44 citizen-age-race/ethnicity categories.[85] The dependent variable was the ACS housing unit-level count for the group, and the explanatory variables were AR counts for each of the 44 groups. The sample was responding 2016-2020 ACS housing units.[86] The coefficients were applied to AR counts in nonresponding ACS housing units to generate estimates for each housing unit-citizen-age-race/ethnicity cell.

The OLS model-estimated shares by age-race/ethnicity for nonresponding units lie between those for responding units, regardless of method/source, and the ones for nonresponding units for all the other methods/sources (Table 3.13E). The OLS estimates are slightly closer to the ones for nonresponding units using the other methods, but they are still quite a bit different from them. The non-Hispanic White shares are much higher, and the Hispanic shares are much lower than the other methods' estimates for nonresponding units, for example.

The OLS model-estimated citizen shares are quite close to those from the ACS using traditional weights (Table 3.13F). For many individual age-race/ethnicity categories, the OLS estimates are between those of the ACS-based estimates and the AR estimates. The main exception is voting-age Hispanics, where the OLS estimate is higher than any of the others, though it is lower than the ACS citizen share estimate for responding housing units. There is very little difference between OLS results with the two different normalizations of the person-place probabilities. In sum, the OLS model method predicts a smaller unit nonresponse bias than the other methods/sources, including the 2020 Census.

As an additional test of the OLS regression method, we estimated 22 age-race/ethnicity models inserting 2020 Census responses in place of ACS responses as the dependent variables, again using the responding ACS housing unit sample. The coefficients were applied to AR for nonresponding ACS housing units. This allows us to see whether the OLS regression predicted values or direct use of AR more closely approximate the estimates from actual 2020 Census responses for nonresponding ACS housing units. Once again, the OLS estimates (Table 3.13G) lie between the 2020 Census estimates in

[85] A drawback of OLS regressions is that they produce negative predicted counts for some cells. The aggregated estimates are positive, though. We also tried Poisson regressions, which produce positive estimates for every cell. The models for some groups did not converge, however.

[86] Housing units classified as vacant or delete by a field representative were included with zero ACS counts.

responding ACS housing units (Table 3.4A, column 1) and nonresponding ones (Table 3.4E, column 1). The 2020 Census estimates in nonresponding ACS housing are closer to the AR estimates (Table 3.4E, column 4) than the OLS estimates for some categories, but not others. Most importantly for citizenship measurement, though, the 2020 Census estimates are much closer to AR estimates than OLS estimates for Hispanics.

Table 3.14 Percent Citizens by Age and Race/Ethnicity, All 2016-2020 ACS Housing Units

| | ACS, EBW | ACS, EBW with AR and Modeled Citizenship in Place of Citizenship Edits and Imputations | Enhanced CVAP, BW | AR, BW | Enhanced CVAP, Population-Adjusted Weights |
|---|---|---|---|---|---|
| Total | 93.34 | 93.09 | 92.61 | 92.04 | 92.56 |
| | Age Categories | | | | |
| Age 0-17 | 97.50 | 97.38 | 97.90 | 98.33 | 97.83 |
| Age 18+ | 92.09 | 91.85 | 90.93 | 89.93 | 90.99 |
| | Age 0-17 | | | | |
| Non-Hispanic | 98.17 | 98.08 | 98.22 | 98.51 | 98.13 |
| NH AIAN Alone | 99.77 | 99.68 | 99.20 | 99.03 | 99.26 |
| NH Asian Alone | 87.26 | 86.92 | 89.30 | 92.96 | 89.17 |
| NH Black Alone | 97.91 | 97.74 | 98.21 | 98.38 | 98.23 |
| NH NHPI Alone | 92.82 | 92.84 | 93.57 | 93.89 | 93.66 |
| NH White Alone | 99.24 | 99.21 | 99.01 | 99.04 | 99.00 |
| NH AIAN & White | 99.95 | 99.93 | 99.72 | 99.28 | 99.70 |
| NH Asian & White | 97.42 | 97.42 | 97.23 | 97.20 | 97.29 |
| NH Black & White | 99.75 | 99.69 | 99.29 | 99.14 | 99.32 |
| NH AIAN and Black | 99.89 | 99.85 | 99.40 | 98.89 | 99.43 |
| NH Other Multi-Race | 99.00 | 99.00 | 98.00 | 97.42 | 98.06 |
| Hispanic | 95.50 | 95.28 | 96.94 | 97.82 | 96.91 |

Table 3.14 Continued

| | ACS | ACS, + AR and Modeled Citizenship in Place of Citizenship Edits and Imputations | Enhanced CVAP, BW | AR, BW | Enhanced CVAP, PEP Population-Adjusted Weights |
|---|---|---|---|---|---|
| | Age 18+ | | | | |
| Non-Hispanic | 95.82 | 95.72 | 95.18 | 94.69 | 95.09 |
| NH AIAN Alone | 99.33 | 99.33 | 97.20 | 96.17 | 97.57 |
| NH Asian Alone | 69.70 | 69.19 | 70.15 | 70.03 | 70.03 |
| NH Black Alone | 95.39 | 95.21 | 95.34 | 95.27 | 95.30 |
| NH NHPI Alone | 82.84 | 82.30 | 80.87 | 80.60 | 80.78 |
| NH White Alone | 98.27 | 98.27 | 97.41 | 96.83 | 97.44 |
| NH AIAN & White | 99.75 | 99.66 | 99.37 | 97.84 | 99.36 |
| NH Asian & White | 91.50 | 91.50 | 90.94 | 90.11 | 91.10 |
| NH Black & White | 97.72 | 97.68 | 97.34 | 96.89 | 97.33 |
| NH AIAN and Black | 99.35 | 99.46 | 98.30 | 96.81 | 98.27 |
| NH Other Multi-Race | 95.31 | 95.25 | 89.78 | 85.67 | 90.25 |
| Hispanic | 72.82 | 71.80 | 69.31 | 67.59 | 69.71 |

Notes: The sample is housing units in the initial 2016-2020 ACS sample. ACS uses traditional person weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS household rosters, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in responding ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding housing units. AR uses AR rosters together with AR and modeled citizenship for both responding and nonresponding housing units. BW is ACS housing unit base weights. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.14 contains citizen share estimates for all ACS housing units using four methods. The first column is ACS citizenship with entropy balance weights. The second column is the same as the first, except replacing ACS edits and imputations with AR/modeled citizenship. Enhanced CVAP (ACS rosters with citizenship like in the second column for responding units, and AR in the nonresponding units) with base weights is in the third column. All AR with base weights is the fourth column. The fifth column is enhanced CVAP with PEP population-adjusted weights.

A comparison of the ACS citizen shares in Tables 3.7A and 3.14 shows that using entropy balance weights instead of base weights lowers the housing unit citizen share estimate from 94.70 to 93.34 percent (1.36 percentage points). In Table 3.14 the citizen share estimate drops a further 0.25 percentage points when replacing ACS edits and imputations with AR/modeled citizenship. It falls 0.48 percentage points if AR with base weights are used for nonresponding ACS housing units in place of entropy balance weights. When AR are used in place of ACS household rosters in responding ACS units, the citizen share estimate declines by 0.57 percentage points. Using population-adjusted weights rather than base weights changes the enhanced CVAP estimates by just 0.05 percentage points (comparing column 5 to column 3), compared to the 1.36 percentage point difference when switching from base weights to final weights for the ACS estimates (comparing column 1 of Table 3.7A to column

1 of Table 3.14). Once again, the differences across the columns are much greater for voting-age Hispanics than for the whole population.

## *3.4 Comparing CVAP Estimates*

In this section we compare the CVAP estimates using the 2016-2020 ACS with EBW person weights to enhanced CVAP, here including both housing units and group quarters.[87] We start with national-level estimates in Tables 3.15-3.16, followed by analysis at lower levels of geography.

Table 3.15 2016-2020 Population by Age and Race/Ethnicity

| | PEP Estimates | ACS, EBW | Enhanced CVAP Before Swapping | Enhanced CVAP After Swapping |
|---|---|---|---|---|
| Total Population | 326,600,000 | 326,600,000 | 326,600,000 | 326,600,000 |
| | Percent by Age | | | |
| Age 0-17 | 22.45 | 22.44 | 22.45 | 22.45 |
| Age 18+ | 77.56 | 77.56 | 77.56 | 77.56 |
| | Percent of Age 0-17 | | | |
| Non-Hispanic | 74.80 | 74.80 | 74.80 | 74.80 |
| NH AIAN Alone | 0.83 | 0.73 | 0.84 | 0.84 |
| NH Asian Alone | 5.26 | 5.01 | 5.25 | 5.25 |
| NH Black Alone | 13.73 | 13.44 | 13.75 | 13.75 |
| NH NHPI Alone | 0.21 | 0.19 | 0.20 | 0.20 |
| NH White Alone | 50.32 | 50.34 | 50.32 | 50.31 |
| NH AIAN & White | 0.45 | 0.55 | 0.46 | 0.46 |
| NH Asian & White | 1.29 | 1.53 | 1.40 | 1.40 |
| NH Black & White | 1.94 | 2.23 | 1.82 | 1.82 |
| NH AIAN and Black | 0.10 | 0.09 | 0.07 | 0.07 |
| NH Other Multi-Race | 0.67 | 0.70 | 0.70 | 0.70 |
| Hispanic | 25.20 | 25.20 | 25.20 | 25.20 |
| | Percent of Age 18+ | | | |
| Non-Hispanic | 83.85 | 83.85 | 83.85 | 83.85 |
| NH AIAN Alone | 0.71 | 0.62 | 0.71 | 0.71 |
| NH Asian Alone | 5.85 | 5.79 | 5.84 | 5.84 |
| NH Black Alone | 12.16 | 12.04 | 12.16 | 12.16 |
| NH NHPI Alone | 0.17 | 0.17 | 0.17 | 0.17 |
| NH White Alone | 63.40 | 63.40 | 63.40 | 63.40 |
| NH AIAN & White | 0.39 | 0.49 | 0.38 | 0.38 |
| NH Asian & White | 0.42 | 0.48 | 0.42 | 0.42 |
| NH Black & White | 0.36 | 0.47 | 0.35 | 0.35 |
| NH AIAN and Black | 0.08 | 0.08 | 0.06 | 0.06 |
| NH Other Multi-Race | 0.31 | 0.32 | 0.35 | 0.35 |
| Hispanic | 16.14 | 16.14 | 16.14 | 16.14 |

Notes: PEP estimates are averages of the Census Bureau Population Estimates Program estimates in 2016-2020. ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS people, as-reported ACS

[87] Here ACS refers to the final 2016-2020 ACS incorporates entropy balance weights in the processing of the 2020 data only. Traditional methods are used for the 2016-2019 data.

citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in group quarters and responding 2016-2020 ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding 2016-2020 housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Although both the ACS with entropy balance weights and enhanced CVAP apply population control adjustments to approximate the PEP estimates, the two versions of the CVAP estimates do not exactly match the PEP numbers due to collapsing smaller age-by-race/ethnicity cells. In Table 3.15 we check how close the two versions are to the PEP at the national level. For enhanced CVAP we include both the unswapped and swapped versions to see what difference swapping makes. The total population numbers are the same to at least four significant digits.[88] For the single race categories and Hispanic, the enhanced CVAP estimates are within 0.02 percentage points of the PEP estimates. Swapping results in differences no larger than 0.01 percentage point. The ACS estimates differ from the PEP by more than the enhanced CVAP estimates do.

Table 3.16 Citizen Shares by Age and Race/Ethnicity for All Housing Units and Group Quarters

| | ACS. EBW | Enhanced CVAP Before Swapping | Enhanced CVAP After Swapping |
|---|---|---|---|
| Total | 93.36 | 92.62 | 92.62 |
| | Age Categories | | |
| Age 0-17 | 97.49 | 97.82 | 97.82 |
| Age 18+ | 92.14 | 91.12 | 91.12 |
| | Age 0-17 | | |
| Non-Hispanic | 98.16 | 98.12 | 98.12 |
| NH AIAN Alone | 99.78 | 99.27 | 99.27 |
| NH Asian Alone | 87.20 | 89.11 | 89.11 |
| NH Black Alone | 97.90 | 98.13 | 98.14 |
| NH NHPI Alone | 92.71 | 93.62 | 93.62 |
| NH White Alone | 99.24 | 99.02 | 99.02 |
| NH AIAN & White | 99.95 | 99.73 | 99.73 |
| NH Asian & White | 97.42 | 97.28 | 97.22 |
| NH Black & White | 99.75 | 99.33 | 99.40 |
| NH AIAN and Black | 99.89 | 99.44 | 99.44 |
| NH Other Multi-Race | 99.00 | 98.06 | 98.04 |
| Hispanic | 95.51 | 96.92 | 96.92 |
| | Age 18+ | | |
| Non-Hispanic | 95.81 | 95.20 | 95.20 |
| NH AIAN Alone | 99.36 | 97.67 | 97.62 |
| NH Asian Alone | 69.60 | 69.93 | 69.93 |
| NH Black Alone | 95.51 | 95.42 | 95.42 |
| NH NHPI Alone | 82.81 | 80.89 | 80.92 |
| NH White Alone | 98.26 | 97.51 | 97.51 |
| NH AIAN & White | 99.76 | 99.39 | 99.39 |
| NH Asian & White | 91.55 | 91.26 | 91.22 |

[88] Census Bureau disclosure avoidance rules do not permit showing more than four significant digits.

| | | | |
|---|---|---|---|
| NH Black & White | 97.74 | 97.39 | 97.41 |
| NH AIAN and Black | 99.41 | 98.48 | 98.48 |
| NH Other Multi-Race | 95.23 | 90.42 | 90.42 |
| Hispanic | 73.07 | 70.07 | 70.02 |

Notes: ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS people, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in group quarters and responding 2016-2020 ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding 2016-2020 housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

As shown in Table 3.16, the citizen share estimate for the total population at the national level is 0.74 percentage points higher in the ACS. Enhanced CVAP has a higher citizen share estimate among children (0.33 percentage points) and lower for the voting-age population (1.02 percentage points). The most consequential differences by race/ethnicity are for voting-age non-Hispanic Whites (0.75 percentage points higher in the ACS) and Hispanics (1.41 percentage points higher for children and 3.00 to 3.05 percentage points lower for the voting-age population with enhanced CVAP). Here swapping changes the enhanced CVAP estimates by up to 0.04 percentage points. The estimates are very similar to those in Table 3.14 for housing units alone (see columns 1 and 5 of Table 3.14), so adding group quarters has little effect on the estimates.

Table 3.17A Mean Percentage Absolute Difference Between County-Level CVAP and PEP Estimates

| | ACS | | Enhanced CVAP Before Swapping | | Enhanced CVAP After Swapping | |
|---|---|---|---|---|---|---|
| | Total | Voting Age | Total | Voting Age | Total | Voting Age |
| Total | 0.31 | 0.61 | 0.95 | 0.92 | 0.95 | 0.91 |
| Non-Hispanic | 0.65 | 0.82 | 1.06 | 1.03 | 1.01 | 0.97 |
| NH AIAN Alone | 65.00 | 64.74 | 22.43 | 22.64 | 19.09 | 18.67 |
| NH Asian Alone | 48.53 | 49.91 | 27.77 | 26.07 | 26.33 | 24.53 |
| NH Black Alone | 29.90 | 30.14 | 15.31 | 15.96 | 14.40 | 15.33 |
| NH NHPI Alone | 120.3 | 119.7 | 80.21 | 82.22 | 77.61 | 80.78 |
| NH White Alone | 0.71 | 0.77 | 1.14 | 1.10 | 1.08 | 1.03 |
| NH AIAN & White | 50.20 | 52.53 | 36.49 | 36.74 | 35.43 | 35.79 |
| NH Asian & White | 78.29 | 88.08 | 55.49 | 68.04 | 53.55 | 67.26 |
| NH Black & White | 58.98 | 83.29 | 35.80 | 56.69 | 34.46 | 55.82 |
| NH AIAN and Black | 130.4 | 132.8 | 117.2 | 120.4 | 116.2 | 119.5 |
| NH Other Multi-Race | 92.16 | 99.20 | 57.17 | 71.06 | 56.00 | 70.67 |
| Hispanic | 8.67 | 11.58 | 4.66 | 5.02 | 4.29 | 4.76 |

Notes: PEP estimates are averages of the Census Bureau Population Estimates Program estimates in 2016-2020. ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS people, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in group quarters and responding 2016-2020 ACS housing units and AR rosters together with AR and modeled citizenship for

nonresponding 2016-2020 housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.17B Weighted Mean Percentage Absolute Difference Between County-Level CVAP and PEP Estimates

| | ACS, EBW | | Enhanced CVAP Before Swapping | | Enhanced CVAP After Swapping | |
|---|---|---|---|---|---|---|
| | Total | Voting Age | Total | Voting Age | Total | Voting Age |
| Total | 0.01 | 0.05 | 0.02 | 0.02 | 0.03 | 0.03 |
| Non-Hispanic | 0.02 | 0.05 | 0.03 | 0.03 | 0.04 | 0.04 |
| NH AIAN Alone | 20.44 | 21.07 | 2.15 | 2.06 | 1.86 | 1.74 |
| NH Asian Alone | 2.80 | 2.07 | 0.31 | 0.26 | 0.30 | 0.26 |
| NH Black Alone | 2.02 | 1.65 | 0.23 | 0.24 | 0.22 | 0.23 |
| NH NHPI Alone | 18.56 | 17.65 | 5.25 | 4.61 | 4.89 | 4.35 |
| NH White Alone | 0.02 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 |
| NH AIAN & White | 27.01 | 27.75 | 17.66 | 16.88 | 17.78 | 17.25 |
| NH Asian & White | 19.77 | 19.16 | 12.33 | 11.67 | 11.39 | 11.31 |
| NH Black & White | 25.18 | 35.16 | 11.16 | 14.64 | 10.79 | 14.41 |
| NH AIAN and Black | 46.39 | 44.86 | 43.52 | 41.59 | 43.20 | 41.51 |
| NH Other Multi-Race | 20.04 | 21.27 | 12.59 | 14.02 | 14.03 | 16.06 |
| Hispanic | 0.09 | 0.15 | 0.08 | 0.09 | 0.09 | 0.00 |

Notes: PEP estimates are averages of the Census Bureau Population Estimates Program estimates in 2016-2020. ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses the people reported in the ACS for group quarters and responding ACS housing units and AR rosters together with demographics from numerous survey and administrative record sources and modeling for nonresponding housing units. All estimates include both housing units and group quarters. The weights for the weighted means are the county-level PEP estimates for the particular cell. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Tables 3.17A-B show average percentage absolute differences between county-level ACS, enhanced CVAP, and PEP estimates.[89] The total and voting-age population differences with PEP are smaller with the ACS, but they are larger for each subcategory except non-Hispanic and non-Hispanic White. The larger ACS absolute differences for subcategories may be due to less collapsing of race/ethnicity categories when constructing population adjustment factors for enhanced CVAP. Swapping slightly reduces the enhanced CVAP differences from PEP (Table 3.17A). The ACS mean absolute differences are much bigger compared to the enhanced CVAP ones when weighting the cells by PEP county-race/ethnicity population (3.17B).

Table 3.18A Median County-Level Coefficient of Variation for Total Population and Voting-Age Population

| | ACS, EBW | Enhanced CVAP Before Swapping | Enhanced CVAP After Swapping |
|---|---|---|---|

[89] The denominator is the average of the two estimates.

| | Total | Voting Age | Total | Voting Age | Total | Voting Age |
|---|---|---|---|---|---|---|
| Total | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| Non-Hispanic | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| NH AIAN Alone | 0.34 | 0.36 | 0.00 | 0.00 | 0.00 | 0.00 |
| NH Asian Alone | 0.18 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 |
| NH Black Alone | 0.09 | 0.08 | 0.00 | 0.00 | 0.01 | 0.00 |
| NH NHPI Alone | 0.74 | 0.76 | 1.15 | 1.14 | 0.92 | 0.94 |
| NH White Alone | 0.00 | 0.00 | 0.22 | 0.19 | 0.01 | 0.01 |
| NH AIAN & White | 0.22 | 0.25 | 0.57 | 0.54 | 0.56 | 0.53 |
| NH Asian & White | 0.42 | 0.52 | 0.85 | 0.95 | 0.84 | 0.97 |
| NH Black & White | 0.32 | 0.49 | 0.59 | 0.83 | 0.54 | 0.82 |
| NH AIAN and Black | 0.71 | 0.72 | 1.22 | 1.23 | 1.22 | 1.25 |
| NH Other Multi-Race | 0.56 | 0.61 | 0.95 | 1.04 | 0.91 | 1.00 |
| Hispanic | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Notes: ACS EBW is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses the people reported in the ACS for group quarters and responding ACS housing units and AR rosters together with demographics from numerous survey and administrative record sources and modeling for nonresponding housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.18B Median County-Level Coefficient of Variation for Citizens and CVAP

| | ACS, EBW | | Enhanced CVAP Before Swapping | | Enhanced CVAP After Swapping | |
|---|---|---|---|---|---|---|
| | Citizens | CVAP | Citizens | CVAP | Citizens | CVAP |
| Total | 0.00 | 0.00 | 0.25 | 0.22 | 0.01 | 0.01 |
| Non-Hispanic | 0.00 | 0.00 | 0.24 | 0.21 | 0.01 | 0.01 |
| NH AIAN Alone | 0.34 | 0.36 | 0.92 | 0.90 | 0.22 | 0.12 |
| NH Asian Alone | 0.27 | 0.30 | 0.52 | 0.49 | 0.46 | 0.47 |
| NH Black Alone | 0.10 | 0.09 | 0.46 | 0.41 | 0.12 | 0.10 |
| NH NHPI Alone | 0.75 | 0.76 | 1.20 | 1.22 | 1.01 | 1.06 |
| NH White Alone | 0.00 | 0.00 | 0.22 | 0.19 | 0.01 | 0.01 |
| NH AIAN & White | 0.22 | 0.25 | 0.57 | 0.54 | 0.56 | 0.53 |
| NH Asian & White | 0.43 | 0.53 | 0.85 | 0.97 | 0.84 | 0.98 |
| NH Black & White | 0.33 | 0.49 | 0.59 | 0.83 | 0.54 | 0.82 |
| NH AIAN and Black | 0.71 | 0.72 | 1.22 | 1.24 | 1.23 | 1.26 |
| NH Other Multi-Race | 0.56 | 0.61 | 0.97 | 1.06 | 0.91 | 1.03 |
| Hispanic | 0.05 | 0.09 | 0.47 | 0.41 | 0.11 | 0.16 |

Notes: ACS EBW is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses the people reported in the ACS for group quarters and responding ACS housing units and AR rosters together with demographics from numerous survey and administrative record sources and modeling for nonresponding housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Table 3.18C Median Block Group-Level Coefficient of Variation for Citizens and CVAP

| | ACS | | Enhanced CVAP Before Swapping | | Enhanced CVAP After Swapping | |
|---|---|---|---|---|---|---|
| | Citizens | CVAP | Citizens | CVAP | Citizens | CVAP |
| Total | 0.18 | 0.17 | 0.34 | 0.30 | 0.24 | 0.23 |
| Non-Hispanic | 0.19 | 0.18 | 0.34 | 0.31 | 0.26 | 0.25 |
| NH AIAN Alone | 0.91 | 0.91 | 1.55 | 1.70 | 1.46 | 1.64 |
| NH Asian Alone | 0.74 | 0.75 | 1.35 | 1.37 | 1.39 | 1.42 |
| NH Black Alone | 0.63 | 0.62 | 1.06 | 1.06 | 1.06 | 1.06 |
| NH NHPI Alone | 0.96 | 0.96 | 1.82 | 1.82 | 1.82 | 1.82 |
| NH White Alone | 0.21 | 0.20 | 0.35 | 0.32 | 0.31 | 0.29 |
| NH AIAN & White | 0.91 | 0.92 | 1.81 | 1.82 | 1.82 | 1.82 |
| NH Asian & White | 0.92 | 0.94 | 1.54 | 1.70 | 1.58 | 1.76 |
| NH Black & White | 0.93 | 0.97 | 1.46 | 1.70 | 1.50 | 1.76 |
| NH AIAN and Black | 0.97 | 0.97 | 1.82 | 1.82 | 1.88 | 1.88 |
| NH Other Multi-Race | 0.96 | 0.97 | 1.58 | 1.70 | 1.59 | 1.76 |
| Hispanic | 0.58 | 0.60 | 0.99 | 1.01 | 0.96 | 1.00 |

Notes: ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS household rosters, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in group quarters and responding ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

The median coefficients of variation of the county- and block group-level estimates are shown in Tables 3.18A-C. Swapping has little effect on the enhanced CVAP coefficients of variation. The ACS median coefficients of variation are smaller in nearly all cells.

## *3.5 ACS vs. Enhanced CVAP Citizen Share by State*

In this section we show state-level voting-age population citizen shares using ACS EBW and enhanced CVAP. Table 3.19 shows that the citizen share differences among the whole voting-age population are generally small, with just one state having a difference of over two percentage points across methods (the ACS has a 2.27 percentage point higher citizen share in California). No state has a difference of over two percentage points for non-Hispanic White alone. There are some large differences in the non-Hispanic Black alone group in states such as South Dakota (10.2 percent lower in ACS), Maine (7.9 percent lower in ACS), and Vermont (6.2 percent lower in the ACS), where that group's share of the population is very low (1.9, 1.1, and 1.1 percent of the voting-age population in South Dakota, Maine, and Vermont, respectively). Similarly, large differences in the non-Hispanic Asian group are found in a few states with low shares for that group (e.g., a 7.47 percent lower citizen share in the ACS in Vermont, which has a voting-age population that is just 1.6 percent non-Hispanic Asian). Among

Hispanics, though, the ACS citizen share is higher in all but three states, and the gap is over two percentage points in 34 states and the District of Columbia.

Figure 3.1 shows that California, Maryland, New York, and Nevada have the largest differences between the overall estimated share of citizens in the voting-age population between the ACS EBW and enhanced CVAP. The differences are highly correlated with the Hispanic share of the voting-age population (0.704 correlation coefficient), which is consistent with the results from Table 3.19 showing large citizen share gaps among Hispanics in most states.

Table 3.19 Percent Citizens in Voting-Age Population in ACS EBW and Enhanced CVAP by State and Race/Ethnicity

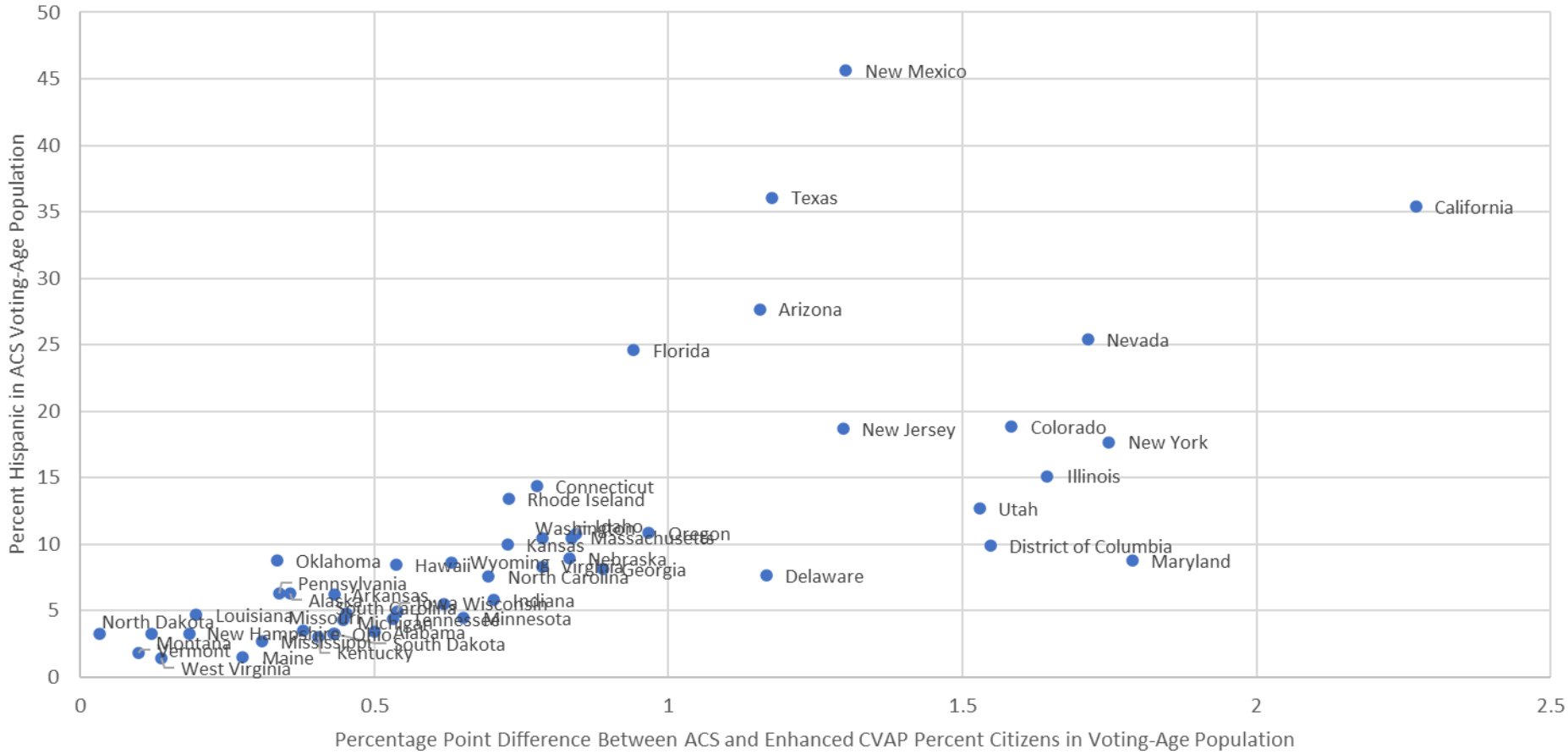
| | Voting-Age Population | | Voting-Age NH White Alone | | Voting-Age NH Black Alone | | Voting-Age NH Asian Alone | | Voting-Age Hispanic | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACS | Enhanced | ACS | Enhanced | ACS | Enhanced | ACS | Enhanced | ACS | Enhanced |
| United States | 92.14 | 91.12 | 98.26 | 97.51 | 95.51 | 95.42 | 69.60 | 69.93 | 73.07 | 70.02 |
| Alabama | 97.61 | 97.11 | 99.53 | 99.07 | 99.48 | 99.39 | 60.43 | 61.66 | 59.73 | 55.32 |
| Alaska | 96.21 | 95.85 | 98.82 | 98.40 | 93.80 | 94.08 | 68.73 | 70.56 | 90.22 | 87.85 |
| Arizona | 91.25 | 90.10 | 97.95 | 97.37 | 94.90 | 94.58 | 64.97 | 67.96 | 78.12 | 75.31 |
| Arkansas | 96.10 | 95.67 | 99.54 | 99.13 | 99.37 | 99.41 | 58.33 | 61.59 | 58.48 | 56.84 |
| California | 84.80 | 82.53 | 96.32 | 94.77 | 96.54 | 96.25 | 75.64 | 75.30 | 73.06 | 68.81 |
| Colorado | 93.85 | 92.27 | 98.77 | 97.94 | 91.63 | 92.24 | 68.58 | 69.02 | 79.61 | 74.35 |
| Connecticut | 92.28 | 91.50 | 97.22 | 96.30 | 90.07 | 89.61 | 62.73 | 64.62 | 78.78 | 78.09 |
| Delaware | 94.72 | 93.56 | 99.20 | 98.08 | 96.45 | 96.06 | 58.96 | 60.76 | 69.19 | 64.32 |
| District of Columbia | 91.68 | 90.14 | 94.92 | 93.24 | 96.93 | 96.36 | 70.96 | 70.12 | 64.38 | 59.02 |
| Florida | 90.00 | 89.06 | 97.14 | 95.81 | 90.95 | 90.41 | 71.15 | 72.42 | 74.74 | 74.53 |
| Georgia | 93.40 | 92.51 | 98.60 | 97.63 | 97.56 | 97.37 | 61.44 | 63.10 | 58.19 | 54.24 |
| Hawaii | 91.32 | 90.79 | 96.94 | 95.23 | 96.79 | 96.27 | 85.88 | 86.03 | 94.42 | 92.88 |
| Idaho | 95.80 | 94.96 | 99.09 | 98.73 | 86.69 | 87.48 | 66.33 | 67.07 | 73.63 | 69.07 |
| Illinois | 91.92 | 90.27 | 97.73 | 96.75 | 97.77 | 97.86 | 67.78 | 67.92 | 70.91 | 64.37 |
| Indiana | 96.25 | 95.55 | 99.45 | 98.87 | 96.49 | 96.98 | 51.52 | 54.87 | 69.15 | 63.36 |
| Iowa | 96.28 | 95.75 | 99.44 | 98.97 | 84.35 | 87.76 | 52.97 | 55.03 | 69.96 | 65.55 |
| Kansas | 95.02 | 94.29 | 99.37 | 98.85 | 96.02 | 96.05 | 64.05 | 65.15 | 68.60 | 65.50 |
| Kentucky | 97.34 | 96.93 | 99.46 | 99.06 | 95.92 | 96.55 | 56.06 | 59.02 | 59.71 | 56.87 |
| Louisiana | 97.22 | 97.03 | 99.36 | 98.94 | 99.45 | 99.46 | 68.52 | 69.10 | 65.15 | 67.01 |
| Maine | 98.26 | 97.98 | 99.12 | 98.83 | 65.48 | 73.38 | 66.90 | 66.92 | 89.94 | 88.64 |
| Maryland | 91.82 | 90.04 | 98.31 | 97.03 | 94.13 | 93.96 | 71.80 | 70.46 | 58.62 | 48.57 |
| Massachusetts | 91.29 | 90.45 | 96.68 | 95.67 | 83.22 | 83.74 | 62.63 | 64.09 | 76.80 | 75.42 |
| Michigan | 96.38 | 95.93 | 98.49 | 98.06 | 98.52 | 98.44 | 59.74 | 61.23 | 78.82 | 75.72 |
| Minnesota | 95.63 | 94.98 | 99.27 | 98.79 | 83.02 | 83.96 | 72.75 | 72.19 | 66.02 | 61.15 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Mississippi | 98.37 | 98.07 | 99.63 | 99.25 | 99.77 | 99.71 | 63.03 | 66.47 | 68.13 | 62.57 |
| Missouri | 97.58 | 97.20 | 99.38 | 99.04 | 97.99 | 98.09 | 61.50 | 63.68 | 76.45 | 72.91 |
| Montana | 98.88 | 98.76 | 99.39 | 99.27 | 89.62 | 94.44 | 69.09 | 71.17 | 93.44 | 92.14 |
| Nebraska | 94.82 | 93.99 | 99.49 | 98.74 | 88.86 | 91.17 | 54.30 | 57.37 | 65.71 | 62.01 |
| Nevada | 88.39 | 86.68 | 97.82 | 96.77 | 96.15 | 96.31 | 77.35 | 76.87 | 68.66 | 64.42 |
| New Hampshire | 96.99 | 96.81 | 98.85 | 98.51 | 82.29 | 83.29 | 61.90 | 64.50 | 81.86 | 82.29 |
| New Jersey | 88.81 | 87.51 | 96.97 | 95.64 | 92.23 | 91.99 | 66.70 | 66.59 | 72.36 | 70.17 |
| New Mexico | 93.62 | 92.32 | 98.89 | 98.19 | 92.85 | 94.61 | 65.37 | 67.41 | 88.71 | 86.30 |
| New York | 89.12 | 87.37 | 96.63 | 95.08 | 89.64 | 88.47 | 66.93 | 65.04 | 75.21 | 72.61 |
| North Carolina | 94.45 | 93.75 | 98.97 | 98.28 | 98.16 | 98.18 | 61.70 | 63.40 | 56.07 | 52.46 |
| North Dakota | 97.00 | 96.56 | 99.25 | 98.69 | 67.19 | 72.31 | 41.34 | 42.12 | 87.07 | 84.31 |
| Ohio | 97.51 | 97.08 | 99.36 | 98.92 | 97.26 | 97.13 | 56.94 | 60.44 | 81.61 | 78.23 |
| Oklahoma | 95.42 | 95.09 | 99.32 | 98.99 | 97.44 | 97.51 | 65.76 | 67.27 | 65.49 | 64.21 |
| Oregon | 93.93 | 92.96 | 98.69 | 98.23 | 91.18 | 92.37 | 69.22 | 70.50 | 69.31 | 64.36 |
| Pennsylvania | 96.34 | 96.01 | 99.03 | 98.55 | 95.53 | 95.90 | 63.71 | 64.90 | 82.87 | 81.36 |
| Rhode Island | 93.27 | 92.54 | 98.14 | 97.37 | 85.77 | 86.10 | 65.01 | 66.17 | 75.89 | 74.61 |
| South Carolina | 96.66 | 96.21 | 99.05 | 98.52 | 99.41 | 99.32 | 65.16 | 66.41 | 59.22 | 57.75 |
| South Dakota | 97.18 | 97.15 | 99.43 | 99.04 | 62.56 | 72.75 | 50.25 | 55.20 | 73.28 | 74.70 |
| Tennessee | 96.54 | 96.01 | 99.25 | 98.78 | 98.50 | 98.49 | 60.68 | 61.88 | 56.20 | 52.47 |
| Texas | 87.44 | 86.26 | 98.21 | 96.77 | 95.62 | 96.03 | 64.93 | 67.17 | 73.99 | 72.19 |
| Utah | 93.66 | 92.13 | 99.04 | 98.20 | 82.89 | 85.70 | 63.11 | 65.58 | 67.19 | 60.01 |
| Vermont | 97.78 | 97.68 | 98.97 | 98.65 | 78.74 | 84.90 | 57.99 | 65.46 | 87.32 | 89.28 |
| Virginia | 93.34 | 92.56 | 98.40 | 97.62 | 96.93 | 96.73 | 70.55 | 70.75 | 64.02 | 61.40 |
| Washington | 91.23 | 90.44 | 97.80 | 97.09 | 87.56 | 89.49 | 66.47 | 68.08 | 67.22 | 63.65 |
| West Virginia | 99.10 | 98.96 | 99.70 | 99.55 | 97.43 | 97.73 | 58.74 | 64.69 | 85.68 | 83.04 |
| Wisconsin | 96.89 | 96.27 | 99.37 | 99.03 | 97.79 | 97.79 | 66.56 | 67.60 | 72.26 | 66.56 |
| Wyoming | 97.49 | 96.86 | 99.32 | 98.96 | 93.42 | 91.93 | 62.33 | 62.34 | 82.62 | 79.80 |

Notes: ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS household rosters, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and

imputations in group quarters and responding ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

Figure 3.1 ACS vs. Enhanced CVAP Percent Citizens (Voting-Age Population)




Notes: ACS is the post-disclosure avoidance 2016-2020 ACS with traditional weights, except for the 2020 portion which incorporates entropy balance weights in the processing. Enhanced CVAP uses ACS people, as-reported ACS citizenship, and AR citizenship and modeled predictions in place of ACS edits and imputations in group quarters and responding 2016-2020 ACS housing units and AR rosters together with AR and modeled citizenship for nonresponding 2016-2020 housing units. All estimates include both housing units and group quarters. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-272).

# 4 Conclusion

This report develops a method using AR to fill in responses for nonresponding housing units rather than adjusting survey weights to account for selection of a subset of nonresponding housing units for follow-up interviews and for nonresponse bias. The method also inserts AR and modeled probabilities in place of edits and imputations for survey citizenship item nonresponses. We produce CVAP tabulations using this enhanced CVAP method and compare them to estimates not using AR at all and ones that use AR to inform nonresponse bias weight adjustments in the 2020 data.

Replacing person weights not informed by AR with weights informed by AR in 2020 (entropy balance weights) changes the estimated 2016-2020 citizen share by just 0.01 of a percentage point, while the enhanced CVAP method produces an estimated 2016-2020 citizen share that is 0.74 percentage points lower than the official ACS estimate that incorporates entropy balance weights in the processing of the 2020 data. The difference in estimates is largest for voting-age Hispanics, who have a 3.05 percentage point lower citizen share with the enhanced CVAP method.

The official 5-year ACS estimates are closer to overall official population estimates from the PEP, which are also used to adjust final weights in the enhanced CVAP. Subpopulation estimates in the enhanced CVAP by age and race/ethnicity are closer to the official population estimates than in the final 5-year ACS, possibly due to less need to collapse race/ethnicity categories when constructing population adjustment factors because the AR data for nonresponding units increased sample sizes in the weighting cells. Enhanced CVAP margins of error are somewhat larger than those published for the official CVAP.

We examine the contributions of the various methodological choices to the overall differences in CVAP estimates. Among people who have both as-reported ACS and AR citizenship, we find very little difference in the citizenship values. The choice of which to use should thus make almost no difference in the resulting estimates. Their high degree of agreement supports using AR citizenship in place of edited or imputed ACS citizenship. It also provides a basis for using AR citizenship in nonresponding housing units.

As previously found by Abowd et al. (2020), we show that citizen shares vary significantly among ACS people for whom the record linkage system could not assign PIKs depending on whether the reason is due to failure to find a link to the Person Identification Verification System (PVS) reference files vs. insufficient PII. Our modeling incorporates this information, while ACS imputations do not. The voting-age Hispanic citizen share estimate is quite sensitive to this choice.

Our analysis finds that modeled predictions and ACS edits, as contrasted with the imputations, perform similarly compared to their respective benchmarks. It is hard to draw strong conclusions, because people with edited ACS citizenship, but with no available AR citizenship make up just 0.03 percent of the sample. The choice to use edits vs. modeled predictions has little impact on these estimates.

We evaluate the estimates produced by AR household rosters, age, and race/ethnicity when used to fill in data for nonresponding ACS housing units by comparing them to those in the 2020 Census, separately for responding and nonresponding ACS housing units. AR produces 2020 total population estimates similar to the 2020 Census responses for the same housing units, especially in nonresponding

ACS housing units. The differences in age and race/ethnicity characteristics across responding and nonresponding ACS housing units are very similar in AR and the 2020 Census, suggesting that AR produce reasonable demographic estimates in nonresponding units.

A decomposition of the difference in the nonresponding housing unit citizen share estimate when using AR to fill in the responses vs. adjusting ACS response weights shows that switching from ACS to AR household rosters and the difference in estimated unit nonresponse bias have similar negative effects on the citizen share estimate. Among voting-age Hispanics, the ACS-to-AR household roster switch dominates, consistent with high omission rates of voting-age Hispanic noncitizens, and especially those with unknown legal status, from responding ACS household rosters. This illustrates how citizenship biases in ACS household rosters makes it hard to address unit nonresponse bias solely through weight adjustments to the ACS people in responding units.

Population control weight adjustments have little effect on the enhanced CVAP estimates, reflecting their similarity to PEP estimates.

Disclosure avoidance protection via swapping has little effect on the coefficients of variation of the enhanced CVAP estimates, even at low levels of geography.

There are several ways in which the enhanced CVAP method can be improved. It currently uses only person identifiers (PIKs) assigned to people with SSNs or ITINs, excluding U.S. residents who have neither. People excluded from the estimates due to not having an SSN or ITIN are mostly noncitizens (Brown et al. [2023]). We could consider using person record linkage incorporating reference files that include people without SSNs or ITINs, following Abowd et al. (2020).

The Abowd et al. (2020) study used some AR sources for noncitizens that were not available for this project. Using such sources would help address holes in coverage of the noncitizen population.

An important source of uncertainty is whether people in the ACS or AR are U.S. residents on the survey reference date. We could study the utility of U.S. arrival and departure data, tax filings and absentee ballot requests from outside the U.S., and expatriate registration at U.S. embassies as sources of foreign residence. Models could be developed to predict the probability that a person is a U.S. resident on the reference date.

Increasing the set of AR sources would not only reduce omissions, but it would also help fill out people's address history as they move. This could reduce errors in placing AR people on a particular date. Improved person-place models could also reduce errors.

The procedure for determining which race/ethnicity values to use when sources conflict and the age and race/ethnicity models could be refined to improve accuracy.

The choice not to alter ACS responses could be revisited. The comparison of ACS to AR household rosters suggests that the ACS may be omitting noncitizens at a higher rate. Besides the option of using AR households in place of ACS households, another option would be to blend the ACS and AR household roster information. For example, if the person-place model predicts that an AR person might be a resident in a household, but (s)he is not in the ACS response, then the AR person could be added to the response with a probability based on the model prediction. Similarly, one might estimate the

probability that an ACS household roster member is actually in the household in the tabulation month, and then weight the household roster member by this probability.

The initial ACS sample is this project's set of housing structures. An alternative to explore would be to include all U.S. residences with AR. Using the full population could improve accuracy at lower levels of geography. The estimates could be timelier, since it wouldn't be necessary to combine multiple years of data.

Improving the enhanced CVAP method would not necessarily cause its estimates to be more like those using the ACS exclusively. We have demonstrated that if we were to make greater use of AR household rosters in the enhanced CVAP method, the estimated citizen share would be lower. Expanding the record linkage beyond SSN- and ITIN-holders and incorporating more AR sources of noncitizens would most likely lower the estimated citizen share even further. Introducing the probability that an AR person is a U.S. resident in the reference period may raise the estimated citizen share, however, so the overall effect of the changes is unclear.

# References


Abowd, John M., William R. Bell, J. David Brown, Michael B. Hawes, Misty L. Heggeness, Andrew D. Keller, Vincent T. Mule Jr., Joseph L. Schafer, Matthew Spence, Lawrence Warren, and Moises Yi (2020) "Determination of the 2020 U.S. Citizen Voting Age Population (CVAP) Using Administrative Records and Statistical Methodology Technical Report," CES Discussion Paper CES 20-33, October 30. Downloaded on April 1, 2022 at https://census.gov/library/working-papers/2020/adrm/CES-WP-20-33.html.

Blevins, Cameron, and Lincoln Mullen (2015) "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction," *DHQ: Digital Humanities Quarterly*, Vol. 9 (3). Downloaded from http://digitalhumanities.org/dhq/vol/9/3/000223/000223.html on April 5, 2022.

Bollinger, Christopher R., Barry T. Hirsh, Charles M. Hokayem, and James P. Ziliak (2021) "Trouble in the Tails? What We know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch," *Journal of Political Economy*, Vol. 127, No. 5.

Brown, J. David, Samuel R. Cohen, Genevieve Denoeux, Suzanne Dorinski, Misty L. Heggeness, Carl Lieberman, Linden McBride, Marta Murray-Close, Hongxun Qin, Allen E. Ross, Danielle H. Sandler, Lawrence Warren, and Moises Yi (2023) "Real-Time 2020 Administrative Record Census Simulation," working paper, U.S. Census Bureau.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi (2019a). "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census," Center for Economic Studies Working Paper Series, No. 18-38R (June). Washington, DC: U.S. Census Bureau.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi (2019b). "Predicting the Effect of Adding a Citizenship Question to the 2020 Census," *Demography*, Vol. 56, pp. 1173-1194.

Brummet, Quentin, Denise Flanagan-Doyle, Joshua Mitchell, and John Voorheis (2018) "Investigating the Use of Administrative Records in the Consumer Expenditure Survey," CARRA Working Papers 2018-01, Center for Economic Studies, U.S. Census Bureau.

Davern, Michael E., Bruce D. Meyer, and Nikolas K. Mittag (2019) "Creating Improved Survey Data Products Using Linked Administrative-Survey Data," *Journal of Survey Statistics and Methodology*, Vol. 7 (3): pp. 440-463.

Ennis, Sharon R., Sonya R. Porter, James M. Noon, and Ellen Zapata (2018) "When race and Hispanic origin reporting are discrepant across administrative records and third party sources: Exploring methods to assign responses. *Statistical Journal of the IAOS*, Vol. 34 (2): pp.179-189. Downloaded on April 1, 2022 at https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji170374

Evans, Sarah, Jenna Levy, Jennifer Miller-Gonzalez, Monica Vines, Anna Sandoval Girón, Gina Walejko, Nancy Bates, and Yazmin Garcia Trejo (2019) "2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Focus Group Final Report," Washington, DC: U.S. Census Bureau, January 24. Downloaded from https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-focus-group.pdf on August 24, 2022.

Fellegi, Ivan P., and Sunter, Alan B. (1969) "A Theory for Record Linkage," *Journal of the American Statistical Association*, Vol. 64 (328): pp. 1183-1210.

Gee, Lisa Christensen, Matthew Gardner, Misha E. Hill, and Meg White (2017) "Undocumented Immigrants' State & Local Tax Contributions," Institute on Taxation & Economic Policy, March. Downloaded from https://itep.sfo2.digitaloceanspaces.com/immigration2017.pdf on March 30, 2022.

Giefer, Katherine, Abby Williams, Gary Benedetto, and Joanna Motro (2015) "Program Confusion in the 2014 SIPP: Using Administrative Records to Correct False Positive SSI Reports," Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference. Downloaded from https://nces.ed.gov/FCSM/pdf/I1_Giefer_2015FCSM.pdf on April 7, 2022.

Jasso, Giullermina, and Mark R. Rosenzweig (2020) "What is the Size and Legal Composition of the US Foreign-Born Population?" *International Migration Review*, Vol. 54 (3): pp. 640-673.

Jensen, Eric B., Renuka Bhaskar, and Melissa Scopilliti (2015) "Demographic Analysis 2010: Estimates of Coverage of the Foreign-Born Population in the American Community Survey," U.S. Census Bureau Population Division Working Paper No. 103, June.

Marks, Rachel, and Merarys Rios-Vargas (2021) "Improvements to the 2020 Census Race and Hispanic Origin Question Designs, Data Processing, and Coding Procedures," Random Samplings, U.S. Census Bureau, August 3. Downloaded from www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to-2020-census-race-hispanic-origin-question-designs.html on June 23, 2022.

Massey, Douglas S., Jorge Durand, and Karen A. Pren (2014) "Border Enforcement and Return Migration by Documented and Undocumented Mexican Immigrants," *Journal of Ethnic and Migration Studies*.

McGeeney, Kyley, Brian Kriz, Shawnna Mullenax, Gina Walejko, Laura Kail, Monica Vines, Nancy Bates, and Yazmin Garcia Trejo (2019) "2020 Census Barriers, Attitudes, and Motivators Study Survey Report," Washington, DC: U.S. Census Bureau, January 24. Downloaded from https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-study-survey.pdf on August 24, 2022.

Meyer, Bruce D., and Nikolas Mittag (2019) "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net," *American Economic Journal: Applied Economics*, Vol. 11 (2): pp. 176-204.

Mulry, Mary H., Tom Mule, Andrew Keller, and Scott Konicki (2021) "Administrative Record Modeling in the 2020 Census," U.S. Census Bureau, April 27. Downloaded from https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf on April 4, 2022.

National Immigration Law Center (2017). "Individual Taxpayer Identification Number (ITIN): A Powerful Tool for Immigrant Taxpapers," January. Downloaded from https://nilc.org/issues/taxes/itinfaq/ on March 30, 2022.

Puckett, Carolyn (2009) "The Story of the Social Security Number," *Social Security Bulletin*, Vol. 69 (2). Downloaded from https://ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html on March 31, 2022.

Rothbaum, Jonathan, Jonathan Eggleston, Adam Bee, Mark Klee, and Brian Mendez-Smith (2021) "Addressing Nonresponse Bias in the American Community Survey During the Pandemic Using Administrative Data," 2021 American Community Survey Research and Evaluation Report Memorandum Series # ACS21-RER-05 and SEHSD Working Paper #2021-24.

Shin, Hyon B. (2021) "An Assessment of the COVID-19 Pandemic's Impact on the 2020 ACS 1-Year Data," *ACS Research and Evaluation Report #ACS21-RER-04*. Downloaded on April 1, 2022 at https://census.gov/content/dam/Census/library/working-papers/2021/acs/2021_CensusBureau_01.pdf.

Tourangeau, Roger and Ting Yan (2007) "Sensitive Questions in Surveys," *Psychological Bulletin*, Vol. 133 (5): pp. 859-883.

U.S. Census Bureau (2013) "U.S. Census Bureau Statistical Quality Standards". Downloaded on April 1, 2022 at https://census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/quality/statistical-quality-standards/Quality_Standards.pdf.

U.S. Census Bureau (2014a) "ACS Design and Methodology Report". Downloaded on April 1, 2022 at https://census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

U.S. Census Bureau (2014b) "American Community Survey Design and Methodology (January 2014) Chapter 11: Weighting and Estimation." Downloaded on April 1, 2022 at https://census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

U.S. Census Bureau (2014c) "American Community Survey Design and Methodology (January 2014) Chapter 12: Variance Estimation." Downloaded on April 1, 2022 at https://census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

U.S. Census Bureau (2019a) “American Community Survey Accuracy of the Data (2019),” Downloaded on April 1, 2022 at https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/ACS_Accuracy_of_Data_2019.pdf.

U.S. Census Bureau (2020a) “American Community Survey. Sample Size and Data Quality,” Downloaded on April 1, 2022 at https://census.gov/acs/www/methodology/sample-size-and-data-quality/U.S. Census Bureau. (2020b) “American Community Survey. Response Rates,” Downloaded on April 1, 2022 at https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/.

U.S. Census Bureau (2022) “American Community Survey Multiyear Accuracy of the Data (5-year 2016-2020).” Downloaded from https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/MultiyearACSAccuracyofData2020.pdf on August 20, 2022.

Van Hook, Jennifer, Frank D. Bean, James D. Bachmeier, and Catherine Tucker (2014) “Recent Trends in Coverage of the Mexican-Born Population of the United States: Results From Applying Multiple Methods Across Time,” *Demography*, Vol. 51, pp. 699-726.

Wagner, Deborah, and Mary Layne (2014) “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research & Applications’ (CARRA) Record Linkage Software,” *Center for Administrative Records Research and Applications (CARRA) Working Paper Series*, #2014-01, U.S. Census Bureau.

Warren, Robert (2019) “In 2019, the US Undocumented Population Continued a Decade-Long Decline and the Foreign-Born Population Neared Zero Growth,” *Journal on Migration and Human Security*, Vol. 9 (1): pp. 31-43.

Zaslavsky, Alan M. (1988) “Representing Local Area Adjustments by Reweighting of Households,” *Survey Methodology*, Vol. 14 (2): pp. 265-288.